OCTOBER
1 9 5 6
VOLUME 20
NUMBER 5

JOURNAL OF CONSULTING PSYCHOLOGY

AMERICAN PSYCHOLOGICAL ASSOCIATION

JOURNAL OF CONSULTING PSYCHOLOGY

Published by the American Psychological Association, Inc.
at Prince and Lemon Sts., Lancaster, Pa.
and 1333 Sixteenth St. N.W., Washington, D.C.

Edited by Laurance F. Shaffer
Teachers College, Columbia University

Advisory Editors: Donald W. Fiske, *University of Chicago*. Ann Magaret Garner, *University of Illinois*. Harrison G. Gough, *University of California*. William A. Hunt, *Northwestern University*. E. Lowell Kelly, *University of Michigan*. Morris Krugman, *Board of Education, New York, N. Y.* Bertha M. Luckey, *Board of Education, Cleveland, Ohio*. Fred McKinney, *University of Missouri*. Anne Roe, *New York, N. Y.* Carl R. Rogers, *University of Chicago*. Nevitt Sanford, *University of California*. Edward Joseph Shoben, Jr., *Teachers College, Columbia University*. George J. Wischner, *University of Pittsburgh*. J. R. Wittenborn, *Rutgers University*.

Lorraine Bouthilet, *Managing Editor*

The *Journal of Consulting Psychology* is the clinical journal of the American Psychological Association. It is devoted primarily to the publication of original research relevant to psychological diagnosis, psychotherapy and counseling, personality, and the dynamics of behavior. Case studies, theoretical contributions, descriptions of clinical techniques, and discussions of the professional problems of clinical psychology also appear in its pages.

Editorial Office

Manuscripts submitted for publication and correspondence on editorial matters should be sent to the Editor, Laurance F. Shaffer, 525 West 120th St., New York 27, N. Y. All manuscripts should be submitted *in duplicate* to facilitate referral to the Advisory Editors. Original figures are prepared for publication; duplicate figures may be photographic or pencil-drawn copies. Authors are cautioned to retain a copy of the manuscript to guard against loss in the mail.

Style of manuscripts. Manuscripts should be prepared in the style described in the Publication Manual of the American Psychological Association (*Psychological Bulletin*, 1952, 49, 389–449). The Manual may be purchased from the Washington office of the Association for $1.00. Otherwise acceptable manuscripts which depart from the standards of the Manual are returned for revision and retyping.

Brief Reports. A one-page condensation of an acceptable research report may be given early publication without charge if the author agrees (*a*) to make 100 mimeographed copies of the full report available to those who request it, (*b*) to send a copy of the full report to the Editor for deposit in the American Documentation Institute, and (*c*) not to have the full report published elsewhere. For detailed instructions on how to prepare a Brief Report, see the back of the *Contents* page of this Journal or write to the Editor.

Early Publication. Authors may secure the early publication of an acceptable article in the first issue to go to press after its acceptance by arranging for the article to be printed as extra pages, increasing the number of pages that subscribers receive. The charge to the author for such early publication is $20.00 per printed page.

Reprints. Authors receive fifty reprints gratis of all articles except early publication articles and Brief Reports. Additional quantities may be ordered when returning proof.

Business Office

Communications concerning subscriptions, change of address, claims for the nonreceipt of a number, advertising, and other business matters, should be sent to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D.C. Address changes must reach the subscription office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

The *Journal of Consulting Psychology* is published bimonthly in February, April, June, August, October, and December, the yearly volume comprising approximately 480 pages. The subscription per year is $8.00, foreign $8.50, single number $1.50.

Entered as second-class matter, March 21, 1955, at the post office at Lancaster, Pa., under the act of March 3, 1879. Acceptance for mailing at special rate of postage provided for in Section 538, Act of February 25, 1925, authorized May 27, 1940.
Copyright © 1956 by the American Psychological Association, Inc.

JOURNAL OF CONSULTING PSYCHOLOGY

October, 1956 Vol. 20, No. 5

Contents

Psychotherapeutics: Success or Failure: *G. R. Pascal and Melvin Zax* - - - - - - - - 325
Relationship of Success and Length in Psychotherapy: *James W. Taylor* - - - - - - - 332
The Representation of Orthopedic Disability in Children's Figure Drawings: *A. B. Silverstein and H. A. Robinson* - 333
A Comparison of Productivity and Fantasy on the Rorschach and Projective Movement Sequences: *William H. Lundin and Boris Breiger* - - - - - - - - - - - - - - - 342
A Continuous Performance Test of Brain Damage: *H. Enger Rosvold, Allan F. Mirsky, Irwin Sarason, Edwin D. Bransome, Jr., and Lloyd H. Beck* - - - - - - - - - - - 343
Stimulus Structure as a Determinant of the Perceptual Response: *M. Powell Lawton* - - 351
Manifest Anxiety and the Palmar Perspiration Index: *Allen D. Calvin, F. J. McGuigan, Sybil Tyrrell, and Marshall Soyars* - 356
Effective Intelligence as Measured by an Unstructured Sentence-Completion Technique: *George E. Copple* - 357
A Note on Porteus Maze and Wechsler-Bellevue Scores as Related to Antisocial Behavior: *Kenneth Purcell* - 361
MMPI Differences Between Parents of Stuttering and Nonstuttering Children: *Leonard D. Goodstein and W. Grant Dahlstrom* - 365
Response Bias and the MMPI: *Eugene H. Barnes* - - - - - - - - - - - - - - - - - 371
Delaying Capacity, Fantasy, and Planning Ability: A Factorial Study of Some Basic Ego Functions: *Jerome L. Singer, Harold Wilensky, and Vivian G. McCraven* - - - - - - 375
The Development of a Short Form of the Manifest Anxiety Scale: *A. W. Bendig* - - - - 384
Perception of Apparent Motion in Organics and Schizophrenics: *Rayford T. Saucer and Herdis L. Deabler* - 385
Emotional and Intellectual Concomitants of Advanced Chronic Alcoholism: *George A. Peters, Jr.* - 390
Repression and Insight as Related to Reaction to Cartoons: *Seymour Epstein and Richard Smith* - 391
The Meaning and Generality of Perceived Actual-Ideal Discrepancies: *Leon H. Levy* - - 396
Learning in the Immediate Post-Electroshock Period: *George A. Zirkle* - - - - - - - - 399
Note on "Changes in Psychoneurotic Patients With and Without Psychotherapy": *Desmond S. Cartwright* - 403
A Short Form of the Wechsler-Bellevue and the Army Classification Battery as Measures of Intelligence: *Donald B. Murphy and Robert D. Langston* - - - - - - - - - - - - 405
"Reliability of the Blacky Test": A Reply to Charen: *Gerald S. Blum* - - - - - - - - 406
A Reply to Blum: *Sol Charen* - 407
Psychological Test Reviews - 408

Brief Reports

The *Journal of Consulting Psychology* will accept Brief Reports of research studies in clinical psychology for early publication without expense to the author. The procedure is intended to permit the publication of soundly designed studies of specialized interest or limited importance which cannot now be accepted because of lack of space. Several pages in each issue will be devoted to Brief Reports, published in the order of their receipt without respect to the dates of receipt of the regular articles. Most Brief Reports appear in the first or second issue to go to press following their final acceptance.

An author who wishes to submit a Brief Report:

1. Sends the Brief Report, limited to one printed page and prepared according to the specifications given below.

2. Also sends to the Editor a full report of the research study, in sufficient detail to give a clear account of its background, procedure, results, and conclusions, which will be filed with the American Documentation Institute to insure indefinite availability.

3. Prepares at least 100 mimeographed copies of the full report, which the author will send without charge to all who request it as long as the supply lasts.

4. Agrees not to submit the full report to another journal of general circulation.

Specifications

Brief Report. The Brief Report should give a clear, condensed summary of the procedure of the study and as full an account of the results as space permits.

To insure that the Brief Report will be no longer than one printed page, its typescript, including all matter except the title and the author's lines, must not exceed 70 lines averaging 42 characters and spaces in length. Set the typewriter margins for short lines of 42 characters, which are 3.5 inches long in elite typing, and 4.2 inches long in pica.

The manuscript of the Brief Report must be *double spaced* throughout. Except for its short lines, it follows the standard style (1). Headings, tables, and references are avoided or, if essential, must be counted in the 70 lines. Each Brief Report must be accompanied by a footnote in the style below, which is typed on a separate sheet and *not* counted in the 70-line quota: [1]

[1] An extended report of this study may be obtained without charge from John Doe, 300 Market St., Prospect 6, Mass. (giving the author's full name and address), or for a fee from the American Documentation Institute. Order Document No. —— from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $—— for microfilm or $—— for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

Extended report. The full report is prepared in the style specified by the *Publication Manual* (1), except that it may be typed with single spacing for economy in photoduplication by the ADI.

Reference

1. American Psychological Association. Council of Editors. Publication manual of the American Psychological Association. *Psychol. Bull.*, 1952, 49, 389–449.

Journal of Consulting Psychology
Vol. 20, No. 5, 1956

Psychotherapeutics: Success or Failure

G. R. Pascal and Melvin Zax[1]

University of Tennessee

Whether or not behavioral change can take place concomitant with procedures instituted in the psychological clinic is a matter of belief. Belief is a subjective matter and the nature of the evidence required to achieve it seems to vary from person to person. Thus, aborigines have been found who do not believe that a relationship exists between sexual behavior and pregnancy in the human female. No amount of statistical significance can convince them of this relationship. The facts do not fit into their theoretical framework. In our culture, there are individuals who believe that a person behaves the way he does because "he is possessed of the devil," or because, for some supernatural reason, he is "meant to be that way." Persons with such a frame of reference will not be convinced by the nature of available evidence bearing on behavioral change in clinical subjects. Other more sophisticated persons reject evidence on statistical grounds, implying that no one has "proved" that behavioral change in clinic subjects is a result of clinical procedures. They are, of course, correct. No one has "proved" it in that sense. The best available published evidence bearing on this matter is that recently put forth by Rogers and Dymond (4) in which behavioral change as a criterion variable is largely neglected. Satisfactory evidence, on statistical grounds, does not now exist. Without an established criterion, without the possibility of efficient controls, it is improbable that a rigorous scientific experiment can be conducted in this area at this time.

[1] The authors wish to express their gratitude to Drs. K. R. Newton, C. N. Sipprelle, and C. H. Swensen, members of the Staff of the Psychological Service Center, who assisted in the preparation of this paper.

Evidence, however, does exist. Belief in it will depend on the frame of reference of the individual. Certain biases can be singled out which seem to be related to the acceptance of the evidence available. One is a conviction that the human organism learns. This bias is one to which all of us will give lip service; lip service because, although we believe in the learning of reading, writing, and arithmetic, we do not, many of us, believe that other habits such as attitudes can be learned. Another one seems to be the conviction that human learning can be controlled, that is, that conditions can be set up which will foster or hinder the learning of specific pieces of behavior. A third might be the belief that the human organism has the capacity to react with what it has learned in one situation to another similar situation. With these beliefs, one is in a fair way to accept the evidence bearing on behavioral change in clinical subjects as a result of or concomitant with certain procedures in the clinical laboratory. The available evidence runs somewhat as follows:

A subject is brought, or comes under his own power, to our clinical laboratory. It is determined, on grounds not relevant here, that the purpose of laboratory procedures is to obtain a decrement in the number of hours the subject spends alone. Whereas the subject now spends, on the average, a total of ten out of seventeen waking hours alone, it is desired that he spend, on the average, less than five hours a day alone. Number of hours spent alone becomes our dependent variable. From a study of the subject's past history and whatever we can learn from psychological tests, we formulate hypotheses concerning the intervening variables related to the dependent variable. Since we cannot observe or change hours spent alone outside the labora-

tory, we need to change behavior in the laboratory that is highly correlated with our dependent variable. Thus, in this particular case, we might set up as intermediate dependent variables such things as number of words uttered by the subject, the amount of time spent by him facing the experimenter, the changes in pitch of his voice (affect), random movements, posture, facial expression, etc. We now set up a controlled stimulus situation in the clinical laboratory, which in interaction with the intervening variables, is calculated to affect the dependent variable.

The behavior in question, number of hours spent alone, can vary in three directions. It can go up, down, or remain stationary. Our *prediction* is that it will go down. From the point of view of our prediction, it can only vary in two directions; it can go down or not go down. Disregarding all other matters except those under immediate discussion, the probability of effecting behavioral change in the required direction is one over two for the single case.

We now institute our laboratory procedures and obtain behavioral change in the predicted direction. We get another case with a different, gross, dependent variable, inability to care for her infant in a young mother, and again obtain behavioral change in the predicted direction. With other cases and other dependent variables we, again and again, obtain gross behavioral change in the laboratory that we "guess" is highly correlated with behavior outside the laboratory. Since these are independent events, we can multiply probabilities. What level of significance will satisfy us, .01, .001, or .0000001?

With this orientation, the writers entered the files of the Psychological Service Center of the University of Tennessee. This Service Center is operated by the Department of Psychology for training and research purposes in connection with the Clinical Psychology Training Program. It takes referrals of children and adults from agencies, physicians, the school systems and hospitals, to mention some of the referral sources. The Service Center takes all cases for diagnosis or treatment which are felt to be within the province of a clinical psychologist working in an outpatient setting.

The Service Center has been in operation in its present form since 1951. During that time there have accumulated 1,022 case files. The files were entered with the aid of a table of random numbers. Prior to entering the files, however, certain rules for selection were set up. All cases were used which met the following criteria: (*a*) any case which was terminated by consent of both therapist and patient, (*b*) any case of long-term, "deep" therapy which was in progress or was interrupted, where the patient was seen in at least one hundred interviews (called Type III therapy), (*c*) any case of supportive brief therapy which was in progress or was interrupted, where the patient was seen in at least thirty sessions (called Type II therapy), (*d*) any case where environmental manipulation was the major goal, where the patient was seen in as many as five interviews (called Type I therapy). These criteria were set up on the basis of considerable previous experience. It was felt that cases which had less than the criterion number of interviews for the type of therapy indicated did not constitute an experiment. The criteria concerning what did or did not constitute an experiment in this context are outside the range of this discussion but are analogous, in the reasoning behind them, to the criteria applied in other more systematic experimental settings. In any case, these rules were set up before entering the files. Cases were excluded in which large environmental effects beyond any possibility of control by the therapist interfered with therapy. In this manner ten cases were selected from the files.

For each case, notation was made, from the case history, of behavior which was considered deviant and it was desired to change, as reported either by the patient or someone else who gave the history, the latter very often being the case with child patients. Thereafter, after conclusion or interruption of therapy, a final summary of the case, progress reports, or statements of the patient were used to make a judgment as to whether or not the behaviors previously noted changed in the predicted direction. No follow-up was used, no attempt was made to find out whether or not the patient maintained the behavior change over a period of time. We hold

no brief for the objectivity of this procedure; we are cognizant of the possible biases of whoever wrote the case records, of the intangibles underlying the patient's "good" reports, etc. The reliability of judgments as to whether or not behavior changed in the predicted direction, is perfect. This finding is understandable because, disregarding the reliability of the case histories themselves, the questions asked of the case records could be answered by a simple yes or no.

Table 1 shows the results obtained, ten hits and no misses. Skepticism seems appropriate for these data. Therefore we pulled out another ten cases meeting the criteria for selection. Table 2 shows the results, eight hits and two misses. Although expansion of the binomial indicated a satisfactory level of sig-

Table 1

First Group of Ten Meeting the Criteria for Selection

File No.	Age	Sex	No. of interviews	Behavior to be changed	Change	Predicted direction
322	30	F	194	1. Fainting spells 2. Overweight	1. None for 2 years 2. Lost 14 pounds	Yes
355	7	M	50	1. Fearful, withdrawing behavior in school	1. Active participation in school activities, improved grades	Yes
155	27	F	70	1. No social contacts with men 2. No contacts with parents	1. Began dating regularly 2. Began visiting parents	Yes
402	28	F	113	1. Unable to care for children and perform household duties 2. Few social contacts	1. Caring for children and doing housework consistently 2. Began regular church attendance	Yes
302	8	M	34	1. Submissive and passive behavior, e.g., beaten by younger children 2. Frequent accidents	1. Began having fights in school and took responsibility for younger sibs 2. None while in treatment or immediately after	Yes
727	51	F	13	1. Uncontrollable weeping 2. Insomnia	1. Ceased completely 2. Sleeping well	Yes
143	35	M	102	1. Multiple phobias 2. Frequent hostile outbursts 3. No dating	1. These subsided almost completely 2. Ceased completely 3. Frequent dating	Yes
69	27	M	135	1. Drinking excessively to relieve anxiety 2. Somatic symptoms, e.g., stomach cramps 3. Frequent prostitutes	1. Not drunk for 10 months 2. None for one year 3. None for one year	Yes
108	26	M	214	1. Infrequent dating 2. Ashamed of poor hearing and refused to wear a hearing aid 3. Few social contacts	1. More dating 2. Bought hearing aid and used it 3. Began to join clubs, etc.	Yes
54	28	M	41	1. Frequent quarrels 2. No plans for future and dissatisfied with job	1. These greatly diminished 2. Entered graduate school with plans	Yes

Table 2

Second Group of Ten Meeting the Criteria for Selection

File No.	Age	Sex	No. of interviews	Behavior to be changed	Change	Predicted direction
380	47	M	10	1. Acute alcoholic episode where he was continually drunk 2. Unable to work	1. Sobered up 2. Working steadily for 2 years	Yes
369	42	M	18	1. Using narcotics 2. Crying spells 3. Hypochondriacal complaints	1. None for 1½ years 2. None for 2 years 3. None for 2 years	Yes
884	5	M	60	1. Unintelligible speech 2. Poor coordination (muscular) 3. Not completely toilet trained 4. Played with electrical outlets and received severe shocks	1. Now easily understood 2. Better coordination 3. Toilet training completed 4. No shocks for 10 months	Yes
384	19	M	162	1. Several attempts at suicide 2. Crying, tremulousness 3. Poor grades at college	1. None for 2½ years 2. None for 2½ years 3. Better grades	Yes
184	32	M	315	1. Took too much insulin for diabetes against doctor's orders 2. Hypoglycemic attacks 3. Frequent awakening at night 4. Very frugal 5. Very passive in dealing with subordinates	1. This corrected 2. None for 3 months 3. Ceased 4. Now buys furniture, etc. and takes trips 5. Has been able to fire subordinates when necessary	Yes
974	32	M	25	1. Had failed doctoral language exam twice 2. Verbally threatened violence toward a professor	1. Passed language exam and prelims later 2. Stopped	Yes
865	21	F	48	1. Few close friends	1. No change	No
430	29	M	40	1. Overt homosexual acts 2. Frequented prostitutes 3. Constant extreme muscular tension	1. No homosexual acts for 8 months 2. None for one year 3. Moderate tension occasionally	Yes
292	38	M	252	1. Inhaled nitrous oxide and induced hallucinations 2. Sat up nights listening to Bible on tape recorder	1. Gradual decrease until total stop 6 months ago 2. Ceased	Yes
171	22	M	31	1. Excessive drinking 2. High blood pressure, somatic weakness, burning sensation in stomach 3. Hyperactive	1. This reduced 2. Became more severe and hospitalization became necessary 3. No change	No

nificance even for this finding (.05) we were nevertheless worried by the trend indicated in the two misses obtained. We selected another ten folders from the files. Table 3 shows the results, ten hits and no misses. The finding satisfied us. We felt we could safely bet on the side of the "hits."

Tables 4 and 5 show all of the cases pulled

from the files, 350. It will be noted that the bulk of these were diagnostic only. The category "Inquiry Only" exists because any first interview conducted at this Center is usually written into the files despite the fact that the inquiring person may never return to the Center. The second largest category listed in the table is that labeled "Too Few Sessions."

Table 3

Third Group of Ten Meeting the Criteria for Selection

File No.	Age	Sex	No. of interviews	Behavior to be changed	Change	Predicted direction
405	8	M	47	1. Failing grades at school 2. Frequent disciplinary problem	1. Began receiving passing grades 2. Marked reduction in misbehavior	Yes
895	24	F	5	1. Indecision over impending engagement 2. Dreams and nightmares 3. Had broken engagements in past and avoided becoming serious with boys	1. Became engaged 2. These ceased 3. Got married	Yes
14	29	F	197	1. Suicide attempt 2. Constant sitting on husband's lap 3. Weeping behavior 4. Few social contacts	1. No further attempts 2. Ceased 3. Ceased 4. Began attending social affairs often	Yes
385	4	M	163	1. No speech 2. No response to others 3. Assumed positions, posturing 4. Habitually constipated	1. No change 2. Attempts to communicate by pointing 3. This done infrequently 4. No problem for 6 months	Yes
616	25	F	44	1. Few social contacts 2. No pleasure in sex relationships 3. Fear of driving a car	1. Joined clubs, etc. 2. Began to achieve orgasm 3. Learned to drive	Yes
67	32	M	93	1. Wife beat him 2. Spells of extreme physical weakness 3. Crying spells	1. This ceased and at times he struck her 2. None for about 3 years 3. None for about 3 years	Yes
110	22	M	25	1. Attacks of nausea and depression 2. Left school and was indecisive about future	1. Ceased for about 3½ years 2. Returned to medical school	Yes
226	20	M	65	1. Failing courses at school 2. Not eating and sleeping well	1. Improved grades and graduated 2. Eating and sleeping well	Yes
170	32	M	155	1. Worked over-long hours 2. Unable to communicate to wife or subordinates at work	1. Ceased 2. Easier verbalizations with people	Yes
325	22	F	77	1. Excessive sleeping 2. Exhaustion and nausea often 3. Hypnogogic experiences 4. No dating	1. This ceased 2. This ceased 3. These remained 4. Got married	Yes

This category is probably the one that would be most suspected in estimating the effects of psychotherapy in our presentation. It will be noted that there are more cases in this category than the thirty that were accepted for the study. Whether or not these can be considered failures in psychotherapy is a matter for discussion beyond the realm of this paper. We feel that the matter is irrelevant to our main purpose. The category "Environment Effect" is broken down and self-explanatory. Of the thirty cases retained for the analysis, the table shows that there were five children and twenty-five adults. Most of the adults were in what we call Type II therapy which consists of attempts to modify variables within and without the person. Type I therapy is simply environmental manipulation. In Type III therapy an attempt is made to concentrate on variables within the individual to effect behavioral change with relatively little attention being paid to the environment.

Obviously, the tables do not "prove" anything. It is hoped, however, that they might contribute to belief and, more importantly, illustrate the possibility of a systematic approach to the problem of success or failure in psychotherapy. It should be pointed out that we have not been concerned with whether the patient was "improved" or "unimproved" as has been done in previous studies (1, 2). It is impossible to see how any judgment can be made, let alone refined statistical procedures applied with unreliable data such as "improved" or "unimproved." Nor have we been concerned with whether or not the mechanisms of ego-defense became more suitable (3) or if the patient was more comfortable or optimistic (4) or a host of other similar measures proposed to estimate the effects of psychotherapy.

Table 4

Analysis of Cases Rejected

Reason for rejection		Number
Diagnostic only		231
Inquiring only		34
Environmental effect		12
Couldn't pay	5	
Hospitalized	3	
Left town	3	
Withdrawn by physician	1	
Too few interviews to meet criterion		43
Total cases rejected		320

Table 5

Analysis of Cases Retained

Type of therapy	Mean No. of Hours Seen	Number	Behavior changed in predicted direction	
			Yes	No
Type I	10	1	1	0
Type II	60	16	14	2
Type III	184	8	8	0
Play	71	5	5	0
Total cases retained		30		

By all means let us make the patient's ego-defenses more suitable. Let us change his attitudes about himself and others. But these have no significance if they are not reflected in the behavior of the patient as it is displayed in the environment in which he has to live. Making ego-defenses more suitable (or what have you) may be an hypothesis concerning a *way* to effect behavioral change but it cannot be used as a criterion. Neither can unvalidated tests and questionnaires (no matter how much they are Q-sorted) be used as criteria. If the ultimate purpose of psychotherapy is to effect behavioral change (and it has to be) then *behavior* must be the criterion. This proposition not only makes sense but it gets us closer to the science of psychology.

The writers are aware that something other than mere behavioral change is considered in the problem of success or failure in psychotherapy. There is, for instance, the consideration of time during which behavioral change is maintained after psychotherapy. The answer to this question pushes us back to our systematic formulation. If behavior is changed by simple environmental manipulation (Type I) then the maintenance of changed behavior will depend on the stability of the changed environment. If, on the other hand, the process of changed behavior was accomplished exclusively by manipulation of variables within the individual (Type III), then the changed

behavior should be maintained with relative freedom from the environment. No thoroughgoing study with results of psychotherapy can neglect a consideration of procedure used to effect behavioral change if the time dimension is to be added to the notion of success or failure. Another consideration is that, although success or failure may be concerned with the achievement, in the environment, of a desired and predicted change in behavior, the question as to what behaviors, and how much they have to be changed, is unknown at this writing. This consideration leads us into the morass known as "psychological health." We shall by-pass this matter and end our polemic by asserting that we believe a person's behavior can be changed in a predicted direction by the systematic use of well-established principles known to the science of psychology.

Received May 28, 1956.
Early Publication.

References

1. Eysenck, H. J. The effects of psychotherapy: an evaluation. *J. consult. Psychol.*, 1952, **16**, 319–324.
2. Landis, C. Statistical evaluation of psychotherapeutic methods. In L. E. Hinsie, *Concepts and problems of psychotherapy*. New York: Columbia Univer. Press, 1937. Pp. 155–169.
3. Miller, J. G. Criteria and measurement of change during psychiatric treatment. *Bull. Menninger Clin.*, 1954, **18**, 130–137.
4. Rogers, C. R., & Dymond, Rosalind F. *Psychotherapy and personality change*. Chicago: Univer. of Chicago Press, 1954.

Journal of Consulting Psychology
Vol. 20, No. 5, 1956

Relationship of Success and Length in Psychotherapy

James W. Taylor

Veterans Administration Hospital, Denver, Colorado

Cartwright (1) has suggested the existence of a "failure zone" between the 13th and 21st interview in client-centered therapy. The present report is made in order to indicate a similar result occurring in a psychoanalytically oriented clinic. In the process of other research, the total number of cases closed in the clinic during a fiscal year (309 cases of varying diagnosis and with different therapists) were evaluated for number of treatment interviews and for results of treatment. Results of treatment are evaluated as, 0 unimproved, 1 mildly, 2 moderately, and 3 greatly improved by the therapist and consultant. While this criterion for improvement leaves much to be desired, such judgments are common in most clinics operating under practical restrictions. Figure 1 combines the data presented by Cartwright in three graphs with the corresponding data of this clinic. Representation of the two scales of improvement on the same figure was not meant to imply their equality; rather, the purpose is to illustrate the similarities of the curves.

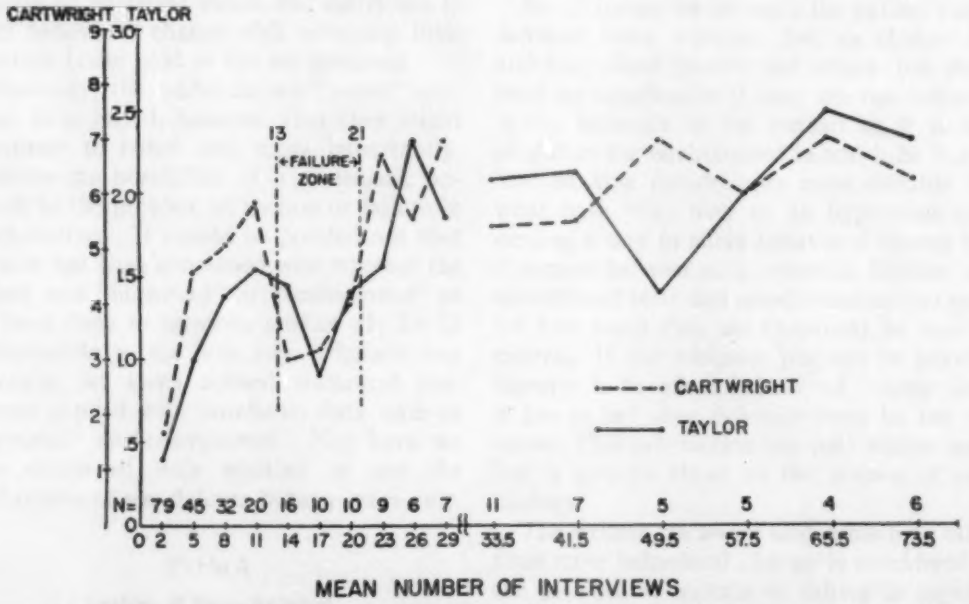
CARTWRIGHT
TAYLOR
13
21
FAILURE ZONE
CARTWRIGHT
TAYLOR
N= 79 45 32 20 16 10 10 9 6 7 11 7 5 5 4 6
0 2 5 8 11 14 17 20 23 26 29 33.5 41.5 49.5 57.5 65.5 73.5
MEAN NUMBER OF INTERVIEWS

Fig. 1. Rated success as a function of number of interviews. The break in the horizontal axis at 30 interviews indicates that both the class intervals of 29 and 33.5 include the 30th interview (4 cases) which was necessary in combining the graphs. *N* indicates the number of patients in each class interval.

Received January 13, 1956.

Reference

1. Cartwright, D. S. Success in psychotherapy as a function of certain actuarial variables. *J. consult. Psychol.*, 1955, **19**, 357–363.

Journal of Consulting Psychology
Vol. 20, No. 5, 1956

The Representation of Orthopedic Disability in Children's Figure Drawings

A. B. Silverstein and H. A. Robinson
Psychiatric Institute, University of Maryland School of Medicine

Twenty years have passed since Schilder first expressed his belief that "the way in which children draw human figures really reflects their knowledge and sensory experience of the body image" (14, p. 106). Since that time many investigators have developed this observation both at the level of theory and in clinical practice. Body image has been variously described by different writers, and the lack of a universally accepted definition has undoubtedly served as a deterrent to systematic research in this problem area. Nonetheless there has remained a widely held belief that drawn human figures do reveal something of the subject's conceptual picture of his own body. Machover, the most active proponent of dynamic interpretations of figure drawings, has stressed that "the process of drawing the human figure is for the subject . . . a problem not only in graphic skill, but one of projecting himself in all of the body meanings and attitudes that have come to be represented in his body image" (10, p. 35). Today, figure drawings are a commonly accepted tool for investigating personality dynamics, self concept, and body image; a familiar part of the clinical battery. This technique, however, shares with many other projective instruments the need for convincing and unequivocal demonstrations of validity.

The present authors are engaged in an investigation of psychological factors in the treatment of and recovery from poliomyelitis,[1] and have had the opportunity to study figure drawings of children in various stages of that illness. In exploratory phases of this work we were frequently impressed with what appeared to be a rather direct reflection of the child's disability in his drawings, and, given impetus by these observations, decided to initiate a more systematic study of the technique. If this procedure could be demonstrated to yield a true picture of body image, it would prove extremely useful, since data from other sources than figure drawings have shown that the discrepancy between body image and the actual physical body is related to adjustment processes in physical disability (2). It should also be noted that figure drawings as such have received scant attention as possible measures of attitudes toward disability. A nonverbal measure of disability-associated attitudes and modes of adjustment would be of great value to workers who deal with the disabled.

The earliest study of the figure drawings of handicapped subjects is reported by Machover (10). Her brief account of this investigation of 20 orthopedic cases provides no methodological details. It would appear that drawings were simply scrutinized for the possible projection of disturbances of body image, and the preliminary results are said to have been striking. Other investigators (1, 6, 12, 16) have taken a similar approach to drawings produced by various types of disabled subjects, and their results, while not as impressive as Machover's, seem to be encouraging. Taken together, these studies suggest that, while relatively few subjects portray their disabilities directly, indirect evidence in the form of distortion and displacement occurs frequently.

[1] Aided by a grant from the National Foundation for Infantile Paralysis.

1. Study of Subjective Classifications

Method. The writers have used a sorting approach in studying a series of drawings by disabled children. Same-sex, opposite-sex, and self-figure drawings were obtained from a group of 22 chronic poliomyelitis cases with paralytic residuals of the lower extremities. Each of the three sets of drawings was evaluated independently by the authors. With no preliminary formulation of a classification scheme, and with negligible disagreement between the judges, we found that these drawings could readily be sorted into six categories. The distinguishing features of these groupings were: (*a*) omission of a portion of the legs; (*b*) external support for the drawn figure, e.g., by an orthopedic appliance; (*c*) marked size discrepancy of the legs; (*d*) improper attachment of the legs to the trunk; and (*e*) unusual treatment of the upper extremities, with the legs judged to be within normal limits. The final category (*f*) included drawings considered to embody neither direct nor indirect evidence of disability.

Results. Four of the 22 self-figures are presented in Fig. 1. Though the drawings differ a great deal, each appears to show clearly and dramatically some representation of impairment. In one way or another these children seem to have projected their physical condition upon their drawn figures. Nor are

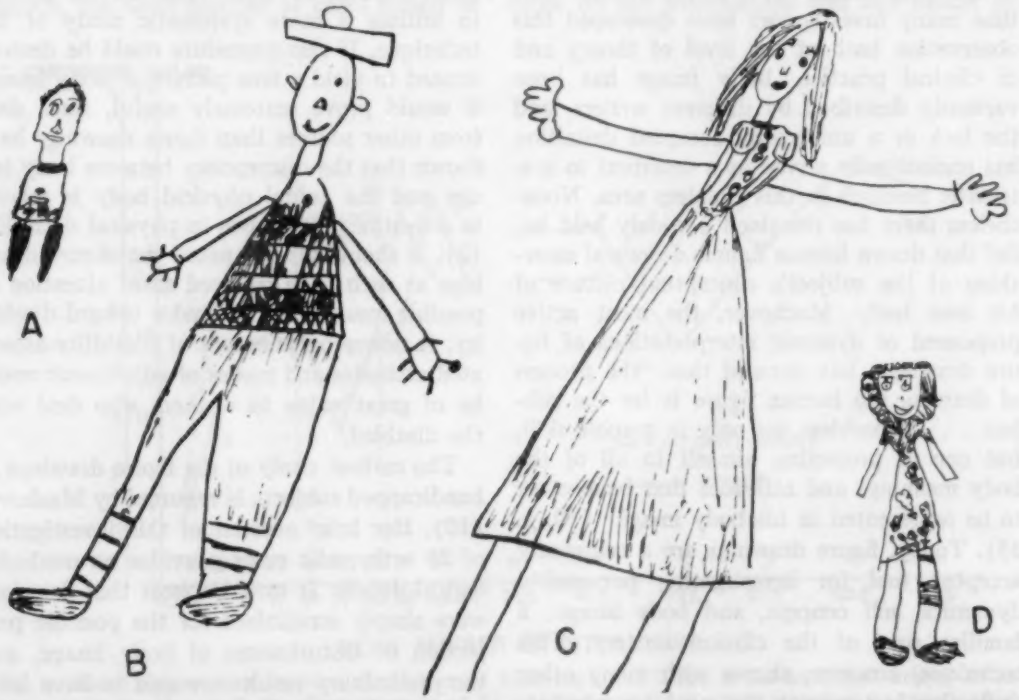

Fig. 1. Self-figures drawn by four children with lower extremity disability following poliomyelitis.

A. Self-figure by a ten-year-old boy in Grade 4. He wears bilateral full-leg braces and has a marked residual weakness of both legs as a result of a spinal poliomyelitis at age two.

B. Drawing by a boy, a nine-year-old second-grader. Disease onset was at age five. Residual is now restricted to the right leg: he wears a short leg brace.

C. Drawing by a nine-year-old girl in the second grade. Since the poliomyelitis onset before her first birthday, she has been readmitted to an orthopedic hospital on five occasions for observation and corrective surgery. Severe paralytic residuals of both legs. She wears a long leg brace with corrective shoes. Note that the legs of her figure are "paper-chopped."

D. Figure by a ten-year-old girl in Grade 3. The disease onset occurred at age 5. Her right leg is essentially flail below the knee, and she wears a long leg brace with pelvic band. This child has had three hospital readmissions for corrective surgery. Note that she has faithfully pictured her brace, though in mirror-fashion.

Table 1

Classification of Figure Drawings by 22 Children With Lower Extremity Disability Following Poliomyelitis

Group	Same-sex	Opposite-sex	Self-figure
Portion of lower extremities omitted	5	3	3
External support provided	0	0	4
Size discrepancy in lower extremities	6	7	5
Improper attachment of legs	3	4	2
Unusual treatment of upper extremities	4	3	4
No evidence of disability	4	5	4
N	22	22	22

these atypical examples; many of the drawings, by direct representation, displacement, distortion, or other unusual treatment, are equally suggestive of body-image disturbance.

Table 1 presents the distribution of figure drawings among the six categories which have been described. In each of the three sets of drawings, more than three-quarters of the subjects were judged to have portrayed a physical disability, either directly or indirectly. By inspection the three sets of drawings show similar distributions among the categories, with the one exception of the "external support" classification. In this instance three self-figures were depicted with orthopedic appliances, and one seated at a desk. This might suggest that the most direct representation of disability may be found in self-figures.

An additional observation should be made with respect to a pattern of consistency found in these drawings. If an individual omits a part of the legs of one of his figures, for example, he tends to treat the other figures similarly. Of the 22 subjects, 16 have either two or three of their drawings in the same category, whereas by chance this should occur only 9 or 10 times. This difference from chance expectancy is significant at the .01 level.

Discussion. While these findings, like those of the investigations cited above, appear very encouraging, it is clear that this subjective approach to the study of figure drawings is seriously deficient from the point of view of scientific rigor. For one thing, the judges who evaluate the drawings are not only aware that they are dealing with productions of handicapped individuals, but also have some general knowledge of the nature and site of the disability. This may introduce bias in at least two ways: the observer may tend to overweigh minor deviations as evidence of disability, or he may de-emphasize certain characteristics not obviously related to the known impairment. Another difficulty in this method of dealing with figure drawings stems from the enormous variety of the productions. Even when working with small samples, one must arbitrarily limit the number of classifications to avoid the unmanageable proliferation of categories. And this inevitably poses the need for an equally arbitrary system of priorities, since many drawings will embody the characteristics specified in more than one category. Drawing A in Fig. 1 illustrates this problem: the figure lacks feet, and the legs are attached to the body in an unusual way. An even stronger objection to this approach is that the frequency of occurrence of the various drawing features in a comparable physically normal sample is unknown. To the extent that the drawings of clinically normal individuals are found to exhibit the same characteristics as those of disabled persons, the utility of the technique is diminished.

Taken together, criticisms of this sort seriously undermine any conclusions which may be drawn from studies employing subjective classification and appraisal of human figure drawings. To be sure, normal controls and objective methods have not been entirely neglected in studying figure drawings of the disabled. Several investigators (3, 12, 15) have compared the drawings of handicapped and normal individuals, using checklists of signs or physical measurements of the drawings. Without exception these investigations yielded negative results, in the sense that drawings of disabled and normal subjects were not reliably distinguished. In each case, however, it would appear that flaws in sampling or methodology may have had the effect of obscuring differences in the drawings. In this connection the following points deserve mention: Failure to control for correspondence

Table 2

Comparison of Disabled and Control Subjects with Respect to Age, Intelligence, and Grade

Group	Age		IQ		Grade	
	Mean	*SD*	Mean	*SD*	Mean	*SD*
Disabled ($N = 22$)	11.2	1.19	94.6	9.27	4.7	1.48
Control ($N = 44$)	11.1	1.02	97.6	7.26	5.2	0.94
t	0.48		1.40		1.52	
F		1.39		1.67		2.54
p	>.10	>.10	>.10	>.10	>.10	<.05

between the sex of the subject and that of his drawn figure denies, in effect, the possibility of differential identification with, and projection on, same-sex and opposite-sex figures. Inclusion of items in scoring schemes which are not directly related to the known impairment, or which cannot in reason be justified as possible displacements or symbolizations of the disability, may needlessly elaborate the measuring instrument without increasing its sensitivity. An example of a seemingly irrelevant item, reported in the literature, is one dealing with the Adam's apple in a study of the drawings of disabled children. Use of extremely heterogeneous groups—a child with a congenital deformity of an arm in the same sample with a child having an osteomyelitis of the leg, for example—seemingly disregards the psychological significance of the nature and site of the disability.

It should fairly be stated that the heterogeneity of disabled individuals poses a difficult and fundamental problem for any research effort in this area. Besides the sources of variation found in physically normal individuals, such considerations as site, degree, and duration of disability make it virtually impossible to secure properly homogeneous samples of adequate size. Thus the research worker must knowingly and selectively sacrifice such variables as seem least pertinent to his goal, and at the same time remain sensitive to the need for homogeneity. The 22 subjects of the present study can be considered homogeneous to this extent: They share the same disease entity and the same general site of disability. They are enrolled in a public school equipped to deal with orthopedically disabled children who by reason of their handicap cannot attend their regular classes. This is an environmental control of a sort, and also ensures that each subject has a marked degree of disability. Age of onset varies greatly in this sample, but each child was at least a year and a half past the acute stage of the disease when he was seen.

2. Controlled Study of Signs

Method. Having considered these subjects in the light of factors related to their disability, it is now necessary to report some other characteristics of the group, and to describe a sample of normal subjects who were introduced as a control. The disabled children ranged in age from 9 to 12 and were equally divided by sex. Over 100 physically normal children from the same metropolitan school system were tested, and 44 of these subjects were selected to form a group comparable to the disabled sample with respect to age, grade, and reported intelligence quotient (Kuhlmann-Anderson). These 44 nondisabled children, equally divided by sex. also ranged in age from 9 to 12. The control group was established at twice the size of the experimental sample to increase the reliability of comparisons, and because physically normal subjects were easily available. The group of children with poliomyelitis residuals, on the other hand, represents nearly all available cases. Table 2 presents a comparison of the two groups.

The equation of the two groups is seen to be entirely acceptable for age and intelligence and mean grade placement; no significant differences are obtained. The significantly greater variance in grade placement observed in the disabled sample may well reflect differing periods of absence from school during treatment for poliomyelitis. With re-

spect to the measures reported, it is submitted that the two groups are essentially comparable.

Each sample was tested in a classroom situation during the school day. The conventional figure-drawing procedure was followed: the subjects were instructed to draw a person, and a person of the opposite sex from the first. They were then asked to draw pictures of themselves. No time limits were set, and no further directions were supplied.

To analyze the drawings obtained, a list of 55 scoring items or signs was developed. Thirty-seven of the signs deal specifically with types of treatment which may be given to lower extremities in figure drawings; for example, shape of foot, length of leg, and omissions. Six signs call for comparisons between the arms and legs of the figures, as in noting differences in the use of shading. The remaining twelve signs involve no consideration of the legs of the figure, but pertain to other features which might be expected to reflect body-image changes resulting from disability, such as the size of the figure and its placement on the paper. Most of these scoring signs were taken directly or adapted from the reports of Buck (5), Goodenough (8), and Machover (10), and the remainder devised by the present writers.

For each sign, specifications were drawn up to increase objectivity of scoring, and a preliminary test of the procedure was satisfactorily completed. The drawings of the experimental and control groups, combined in a random order, were then scored independently by the authors. Each sign was scored plus, minus, or unscorable. As an example, a figure with misshapen feet would receive a plus score for a particular item; a figure with feet judged to be normal would be scored minus; and a figure from which the feet had been omitted would be considered unscorable for this item. Of the 10,890 individual signs (55 signs for each of three drawings by 66 subjects), 88 per cent were scored in agreement by the two judges. Scoring disagreements were later resolved in conference. For individual signs, the percentage of initial agreements ranged from 54 to 98 per cent, but for only three items was it less than 80 per cent. An agreement of 33 per cent would be expected by chance. On the whole, the obtained scoring reliability is considered to be acceptable.

Results. In statistical analysis, the three sets of drawings (same-sex, opposite-sex, and self-figures) were treated separately. For each set, chi square corrected for continuity was employed to compare the drawings of the two subject groups on the 55 signs. Fisher's method (7, p. 96) was used to determine the exact probability associated with chi squares which approached significance. For each comparison, the null hypothesis was tested against the alternative that the disabled group contained a greater proportion of subjects scoring plus on a given sign. Table 3 presents a summary of the statistically significant differences found among the 165 tested.

In the 55 comparisons made for each of the three sets of drawings, two or three dif-

Table 3

p Values Exceeding the Five Per Cent Level in Comparison of Figure Drawings of Disabled and Control Subjects

(Total number of comparisons for each figure is fifty-five)

Item	Same-sex	Opposite-sex	Self-figure
3. Figure placed at side edge of paper			.039
16. Horizontal midline (waist area) emphasized	.017		
20. Feet differ in size or shape (front view)		.035	
22. Leg drawn lighter, thinner, or fainter than corresponding arm		.049	
23. Figure placed above horizontal midline of paper	.014		
26. Length of leg greater than twice length of trunk		.037	
28. One leg thinner than the other			.029
51. Width of leg increases toward foot		.039	
55. Leg attached to but not continuous with body		.0002	.009

ferences significant at the .05 level or better might be expected on the basis of chance alone. It would be necessary to obtain at least six such differences to rule out the operation of chance factors (at the .05 level) (13). The number of differences actually found to be significant at the .05 level or better, for the same-sex, opposite-sex, and self-figures, was two, five, and three respectively.[2] It is accordingly clear that the drawings of the disabled and normal subjects have not been reliably distinguished by the scoring system employed.

Discussion. The results of this objective and analytic approach are plainly less encouraging than those obtained when the drawings, as described earlier, were sorted into a limited number of categories. It happens that two drawing features are given similar consideration in these widely different methods of appraisal, and it is of some passing interest to scrutinize them. In the initial sorting procedure, malattachment of legs to body was thought to be a rather direct representation of impairment. Among the 55 signs there appears a virtually identical item (No. 55). The sign analysis does in fact show that—for two of the three figures—the disabled group contains a significantly greater proportion of subjects scoring plus on this item. In the other instance, the omission of a part of the leg placed a drawing in one of the original categories, and again this item is duplicated in the list of signs. Here the sign approach does not appear to bear out the original impression, namely, that omitting a lower extremity part is an evidence of disability. In their drawings the handicapped children do not more frequently omit parts of legs than do the controls. On the contrary, a significantly greater proportion of the normal subjects makes such omissions. (The difference, for which p is less than .05, lies in the unpredicted direction, and is therefore not reported in Table 3.) This, clearly, is an example of a finding of a subjective and impressionistic approach which does not stand up under more careful scrutiny.

From the methodological point of view, any subjective classification method is clearly inferior to the objective and analytic sign approach. It should, however, be noted that this latter method is itself not beyond criticism. As mentioned earlier, the bulk of signs used in this study dealt specifically with types of special treatment given to the legs of the drawn figures, with the rationale that the primary disease residual was in the lower extremities of the subjects. Yet it may be that restricting the choice of signs in this way has unduly de-emphasized other distinctive treatments, as when the body-image representation of the physical disability generalizes, or is displaced to other portions of the body than the legs. At a somewhat different level, the criticism has been offered that "the checklist [of signs] violates that holistic, dynamically oriented approach which competent clinicians use in making drawing analyses" (4, p. 179). Criticisms of this kind seem entirely reasonable, and must be taken seriously. It is possible, with the materials of the present study, to utilize a more global, "clinical" approach and at the same time retain a measure of scientific rigor.

Clinical judgments as to whether figure drawings were executed by disabled or normal subjects have been reported in the literature (3, 15). The findings of these studies have been negative, i.e., the drawings of normal and disabled groups were not reliably distinguished. However, some of the same flaws (heterogeneous samples and the failure to control for sex) found in studies involving the sign approach may have also served to obscure differences between groups in investigations using clinical judgments. The relative homogeneity of subjects in the present study, and control of the sex factor, here provide a better opportunity for such differences to appear.

3. Controlled Study of Global Judgments

Method. Each of three judges was given the three randomized sets of same-sex, opposite-sex, and self-figure drawings, with instructions to select—by whatever cues the drawings

[2] These figures refer only to differences significant in the predicted direction. Three differences which were statistically significant in the unpredicted direction—where the control group exceeded the disabled sample in proportion of plus scores—are not reported in Table 3.

might provide—the 22 figures in each set which had been produced by the physically disabled children. The three judges were postdoctoral clinical psychologists experienced in the psychodiagnostic use of figure drawings. Before evaluating the drawings, they were given information about the general characteristics of the two subject groups (age, sex, and IQ), and about the special characteristics of the chronic poliomyelitis children (nature, site, and duration of disability). For an individual figure, the judges knew only the sex of the child who had made the drawing; their task, in effect, was to choose—from 66 drawings—11 drawings by boys and 11 by girls which they considered to be productions of disabled subjects.

Results and discussion. The choices of each of the three judges were first compared with those of the others, for the three sets of drawings. On the basis of chance, 56 per cent of the judgments of any two clinicians would be expected to agree. (Two judges, who agree by chance on one-third of their choices of disabled children's drawings, will also have implicit agreement on two-thirds of their evaluations of figures by normal subjects.) The actual percentage of agreements between judges ranged from 52 to 79 per cent for the nine possible combinations (three sets of drawings seen by three judges). Four of the comparisons yielded better than chance agreement (.05 level), suggesting that to some extent the judges were using similar cues.

In considering the ability of the judges to distinguish between the drawings of disabled and normal children, 56 per cent of the selections might again be expected to be correct by chance. Correct judgments actually ranged from 52 to 64 per cent for the nine sets of evaluations. No one set of drawings was more accurately appraised than any other, and no one judge was superior to the others in his ability to distinguish among the drawings. Nor were any judge's ratings significantly better than chance.

These experimental findings, like those obtained with the sign approach, do little to enhance the value of the figure-drawing technique for the study of body image in physical disability. The promise of the original impressionistic sorting method has been fulfilled neither by an objective, analytic procedure, nor by a controlled holistic approach.

Discussion and Implications

A number of hypotheses may be suggested to account for the negative results which have been reported, though none of these hypotheses can be tested with the present data. Harris and Goodenough, on the basis of a large-scale investigation of children's drawings, have concluded that it is somewhat doubtful whether children actually do project the body image upon drawn figures. They remark that "so much of the drawing technique in children under 12 or 13 years of age relates to concept formation and understanding, that probably only a small part of drawing behavior relates to the child's interest, temperamental qualities, and other personality features." [3] Since few of the subjects in the present study were older than twelve, it may be that only a small part of the drawing behavior sampled is related to body image.

Levy's exposition of the technique contains some pertinent observations. Noting that "some clinicians interpret each drawing as a projection of the body image" (9, p. 260), he acknowledges that this sort of projection does frequently occur. He nonetheless goes on to point out that a drawing may also represent many other things, e.g., "a projection of attitudes toward someone else in the environment, a projection of ideal self-image, or a result of external circumstances." Levy remarks, moreover, that "it is usually a combination of all of these." Little wonder that the body image is so difficult to tease out of so complex and intricate a production!

A third possible explanation for the negative results stems from an observation by Machover, who writes that "a leg or an arm may be omitted or distorted for reasons of crippling anxieties, just as surely as for the physical crippling of those limbs" (11, p. 263). This is a powerful and compelling argument: if psychological and physical factors are reflected equally and indistinguishably in drawings, any attempt to differentiate the drawings of disabled and normal subjects

[3] Harris, D. B., personal communication, January, 1955.

seems doomed to failure, however rigorously controls are established. Having made this attempt with particular and systematic attention to problems of control, sampling, and methodology, and without securing positive results, we are now inclined toward this pessimistic point of view. The present study and other similar investigations have rested on the assumption that the physical body, the body image, and the drawn figure are in isomorphic relation. Given careful exploration, this one-to-one relationship does not seem to exist.

We are not prepared, on the other hand, to concede that the figure-drawing technique is without potential value in the study of physically disabled persons. A more individualized approach, focusing on the adjustment patterns of each subject, may bring greater understanding of the widely varying modes of expression displayed by poliomyelitis children in their drawings. For example, to consider those children whose drawings include a clear depiction of their orthopedic appliance—are they, as Machover suggests, "basically body-centered individuals whose whole lives have been blighted by their handicap?" (11, p. 262). Or are they children who have so completely accepted their handicap and incorporated it in their body image, that they feel free to represent it in their drawings? In what ways do such children differ from those who completely omit the legs in their drawings, and from those whose drawings show no evidence whatever of physical disability? In future research designed to answer such questions, we propose systematically to explore these individual differences in terms of psychological processes relevant to physical disability, e.g., defense mechanisms. We are, finally, in complete agreement with Machover's observation that "organs are set in personalities, and have no disconnected life of their own, and . . . the total reaction to handicap . . . is a more vital measure than the specific concrete projection" (11, p. 264).

Summary

Sets of human figure drawings—same-sex, opposite-sex, and self-figures—were obtained from 22 children in the chronic stage of poliomyelitis. These drawings were studied by three different methods to determine whether they reflected the physical disability of the subjects.

1. By inspection, more than three-quarters of the subjects appeared to represent their disability, either directly or indirectly, in their drawings.

2. Comparison of drawings of the disabled children with those of normal controls, equated for age, sex, and IQ, showed a non-significant number of significant differences between the two groups on a series of 55 scoring signs.

3. Judges experienced in the psychodiagnostic use of figure drawings were unable to differentiate the drawings of disabled and normal subjects at a level better than chance.

Various hypotheses were suggested to account for the negative results, and a weakness in the basic assumption underlying the present investigation and similar studies was noted and discussed. Plans for future research, involving a more individualized and personality-centered approach to the human figure-drawing technique, were outlined.

Received January 6, 1956.

References

1. Abel, Theodora M. Figure drawings and facial disfigurement. *Amer. J. Orthopsychiat.*, 1953, 23, 253–261.
2. Bender, Lauretta. Psychoses associated with somatic diseases that distort the body structure. *Arch. Neurol. Psychiat.*, 1934, 32, 1000–1024.
3. Blanchard, Helen M. A preliminary study of the drawing of the human figure by brain injured children. Unpublished master's thesis, Univer. of Denver, 1952.
4. Brown, F. House-Tree-Person and human figure drawings. In D. Brower & L. E. Abt (Eds.), *Progress in clinical psychology*. Vol. I, Section 1. New York: Grune & Stratton, 1952. Pp. 173–184.
5. Buck, J. N. The H-T-P technique. *J. clin. Psychol. Monogr. Suppl.*, 1948, No. 5.
6. Fielding, B. B. Attitudes and aspects of adjustment of the orthopedically handicapped woman. Unpublished doctor's dissertation, Teachers Coll., Columbia Univer., 1950.
7. Fisher, R. A. *Statistical methods for research workers*. (11th Ed.) New York: Hafner, 1950.
8. Goodenough, Florence L. *Measurement of intelligence by drawings*. Yonkers, N. Y.: World Book Co., 1926.

9. Levy, S. Figure drawing as a projective test. In L. E. Abt & L. Bellak (Eds.), *Projective psychology*. New York: Knopf, 1950. Pp. 257–297.
10. Machover, Karen. *Personality projection in the drawing of the human figure*. Springfield, Ill.: Charles C Thomas, 1949.
11. Machover, Karen. Discussion of Theodora M. Abel, "Figure drawings and facial disfigurement." *Amer. J. Orthopsychiat.*, 1953, **23**, 262–264.
12. Prater, G. A. A comparison of the head and body size in the drawing of the human figure by hemiplegic and non-hemiplegic persons. Unpublished master's thesis, Univer. of Kentucky, 1950.
13. Sakoda, J. M., Cohen, B. H., & Beall, G. Test of significance for a series of statistical tests. *Psychol. Bull.*, 1954, **51**, 172–175.
14. Schilder, P. *The image and appearance of the human body*. New York: International Universities Press, 1950.
15. Simms, Nancy. An analysis of the human figure drawings of orthopedic and non-orthopedic children. Unpublished master's thesis, Univer. of Nebraska, 1951.
16. Wille, W. S. Figure drawings in amputees. *Psychiat. Quart. Suppl.*, 1954, **28**, 192–197.

Journal of Consulting Psychology
Vol. 20, No. 5, 1956

A Comparison of Productivity and Fantasy on the Rorschach and Projective Movement Sequences[1,2]

William H. Lundin and Boris Breiger
Chicago State Hospital

Clinicians frequently rely upon inference when analyzing sparse or unelaborated projective records. Clinical utility increases when such records more clearly show psychodynamic content. Experience with the Projective Movement Sequences Test (PMS), a brief motion picture comprising eight semistructured and ambiguous sequences, suggests that psychodynamic content can be increased by providing *Ss* with a moving stimulus (1).

This assumption was tested by using criteria of productivity and fantasy equally relevant to the Rorschach and PMS. Criteria of productivity were: (*a*) number of cards or sequences rejected, (*b*) number of human movement responses, and (*c*) number of animal movement responses. Total *R* was not used as a criterion since Rorschach and PMS response units are not strictly comparable. Structurally, PMS protocols more closely resemble TAT stories than Rorschach records. Three measures of fantasy production were obtained: (*a*) total number of unconscious fantasies, indicated by archaic, bizarre, or highly personalized content, further categorized into (*b*) static objects or fragments of personalized content, i.e., "womb," "broken arm," "dead person," and (*c*) descriptions of active elaborated processes, i.e., "embryo being born," "arms and legs falling off," "cancer inside my body eating away."

The sample consisted of 40 schizophrenics who were given both tests. Records were analyzed by the sign-rank test, showing results significant beyond the .01 level in favor of the PMS on the following criteria: rejection, human movement, animal movement. Total number of personalized fantasies did not differ significantly; however, number of "active elaborated processes" was significantly greater on the PMS (.01), and number of "static objects" was significantly greater on the Rorschach (.01). In the sample, 32.5% of *Ss* rejected more Rorschach cards than PMS sequences, 60% saw more human movement on the PMS, 75% saw more animal movement on the PMS, 62.5% had more personalized fantasies on the PMS, 82.5% had more static fantasies on the Rorschach, and 90% saw a greater number of elaborated fantasies on the PMS. Those *Ss* who rejected Rorschach cards responded to every PMS sequence.

Several hypotheses are suggested: (*a*) stimuli in motion lower the threshold for *M* type responses, (*b*) many times the Rorschach reflects the *S*'s affective dissociation and gives no clear evidence for the unconscious fantasies underlying his behavior, (*c*) for most of these *Ss* the PMS seems to give the necessary "push" towards elaborating fantasies and revealing affect which the clinician is then able to interpret in an unambiguous manner.

Brief Report.
Received May 11, 1956.

[1] Based on a paper read at the 1956 meeting of the Midwest. Psychol. Ass., St. Louis, Mo.

[2] An extended report of this study may be obtained without charge from William H. Lundin, Psychology Dept., Chicago State Hospital, Chicago 34, Ill., or for a fee from the American Documentation Institute. Order Document No. 4972 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

Reference

1. Lundin, William H. Projective movement sequences in the study of personality. *J. proj. Tech.*, 1954, **18**, 208–220.

Journal of Consulting Psychology
Vol. 20, No. 5, 1956

A Continuous Performance Test of Brain Damage[1,2]

H. Enger Rosvold,[3] Allan F. Mirsky,[3] Irwin Sarason,[4]
Edwin D. Bransome, Jr.,[5] and Lloyd H. Beck[6]
Yale University

This paper presents the results of an investigation using a new instrument for the study of brain damage in human subjects. The design of the instrument, the Continuous Performance Test (CPT) was based on certain electroencephalographic evidence which suggested that brain-damaged individuals should show inferior ability as compared with non-brain-damaged individuals on tasks requiring sustained attention or alertness.

The waking EEGs of brain-damaged patients generally show either random bursts of hypersynchronous (high amplitude) activity intruding upon the normal activity of the brain, or a general hypersynchrony (3, 8). Hypersynchronous activity is also evident in the recording from the brain of a sleeping subject (3). If hypersynchrony is associated with reduced vigilance or attention, as suggested by its presence during sleep, then the hypersynchrony of the brain-damaged patients might also indicate reduced attention. For example, according to Hebb (1), hypersynchrony might interfere with the sequential firing of cell assemblies and thereby disrupt the process of attention. Other research strongly suggests that some such relationship does in fact exist. Thus, on the basis of a study of 75 prefrontal lobotomy cases, Levin states:

> Immediately following operation the patient usually lapsed into a drowsy-anergic state from which he could be temporarily aroused by stimuli, only to lapse once again into somnolence when the stimuli were removed. During this drowsy-anergic state EEGs showed diffuse slow activity and a good deal of baseline oscillations of ½ to 2 per second anteriorly mixed with slow rhythms of 2 to 6 per second frequency. As the patient emerged from the drowsy stage the preoperative pattern usually asserted itself (2, p. 422).

The usual measures of attention or alertness such as the digit span and digit symbol substitution subtests of the Wechsler-Bellevue have not consistently showed decline following brain damage (4, 5, 6, 7). In the case of the patient who shows intermittent bursts of hypersynchronous activity, and hence only momentary lapses in attention, the test performance may be due to the fact that the *S* can to a great extent choose his own time to respond, and may reorganize his attention between momentary lapses. The lapses would then not affect his score to any measurable extent.

In the case of the individual who shows a generalized hypersynchrony of the electrical

[1] The authors would like to extend their appreciation to the following individuals and institutions for providing advice and/or subjects and testing facilities for this research: Mr. Samuel Greenhouse of the National Institute of Mental Health, Bethesda, Maryland; Dr. C. Edward Stull and the Southbury Training School, Southbury, Connecticut; Drs. Walter Landmesser and Russell Fuldner and the Newington Crippled Children's Hospital, Newington, Connecticut; Dr. Rhoades and the New Haven Cerebral Palsy Clinic, New Haven, Connecticut; Drs. Milton Senn, William German, Janice Stevens, Morris Wessell, Ethelyn Klatskin, and the staff of the several pediatric and neurological clinics of the Grace-New Haven Community Hospital, New Haven, Connecticut.

[2] Supported in part by grants to H. Enger Rosvold from the Veteran's Administration, Contract VA1001–M3222 and the National Science Foundation.

[3] Now at the National Institute of Mental Health, Bethesda, Maryland.

[4] Now at the West Haven V.A. Hospital, West Haven, Connecticut.

[5] Now at the College of Physicians and Surgeons, Columbia University, New York, N. Y.

[6] Now at the University of Michigan, Ann Arbor, Michigan.

activity of the brain, the classical tests might not reflect his deficiency because they do not require sustained attention over a sufficiently long period of time.

These considerations suggested that a test which would not allow the patient to choose his own time to respond, and which would require a high level of continuous attention over an appreciable interval of time might reflect a deficit that other procedures would miss. Accordingly, the CPT was designed to provide two attention tasks, labeled *X* and *AX*, with *AX* designed to be the more difficult. This paper presents data on the CPT performance of comparable groups of brain-damaged and non-brain-damaged individuals.

If the research demonstrates that there are indeed differences between brain-damaged and non-brain-damaged groups on this test, one of the questions which arises is whether or not the differences are large enough to be of diagnostic use. The data also provide a preliminary answer to this question.

Method

Subjects. No attempt was made in this study to select patients on the basis of either locus, extent, or type of brain damage. The term "brain-damaged" is here used to include any individual for whom there was medical evidence (surgical notes, neurological examination, EEG) of brain pathology. Three groups were studied, each consisting of a brain-damaged and a non-brain-damaged subgroup. The *Defective* group consisted of 72 institutionalized feebleminded *Ss*, who were diagnosed as either of "organic etiology" (the *Brain-Damaged-Defective* subgroup) or of "familial or idiopathic etiology" (the *Non-Brain-Damaged-Defective* subgroup). The *Child* group comprised 45 children from a children's hospital and from various pediatric clinics who were being treated for brain disorders such as cerebral palsy (the *Brain-Damaged-Child* subgroup) or for non-nervous-system disorders (the *Non-Brain-Damaged-Child* subgroup). The third group, designated the Adult group, consisted of 50 adult *Ss*. Half of these (the *Brain-Damaged-Adult* subgroup) were either epileptics being treated at a seizure clinic or brain surgery patients being seen at a neurosurgery clinic. The remaining 24 *Ss* (the *Non-Brain-Damaged-Adult* subgroup) were either part of the general medical population of a hospital or were nonpatient *Ss*.

The age and a measure of intelligence were obtained for all *Ss*. In addition, the length of institutionalization was obtained for the Defective group. The *Ss* in the Defective group were routinely tested on the Stanford-Binet upon admission to the training school; all other *Ss* were tested for IQ just prior to presenting them with the CPT. The intelligence measure used for the Child group was the Stanford-Binet vocabulary, while the *Ss* in the Adult group were tested on either the Full Scale Wechsler-Bellevue, the Verbal Scale of the Wechsler-Bellevue, or the 20-minute modification of the Otis Self-administering Test of Mental Ability, Form B. Approximately equal numbers of *Ss* in each of the Adult subgroups were tested on each of the three measures of intelligence. The average ages and IQs of the several groups are presented in Table 1. The mean age of the Non-Brain-Damaged-Child subgroup was significantly higher than that of the Brain-Damaged-

Table 1

Mean Ages and IQs of the Subjects

Group	*N*	Mean age	*SD* age	Mean IQ	*SD* IQ
Defectives					
Brain-damaged	29	27.3	8.1	63.3	7.8
Non-brain-damaged	43	25.1	7.8	66.8	7.5
Children					
Brain-damaged	19	9.0**	3.0	101.9	28.2
Non-brain-damaged	26	12.3**	2.9	104.6	21.2
Adults					
Brain-damaged	25	31.7	12.1	90.5*	15.2
Non-brain-damaged	25	31.5	13.9	100.0*	15.0

* Indicates a significant difference ($p < .05$).
** Indicates a significant difference ($p < .01$).

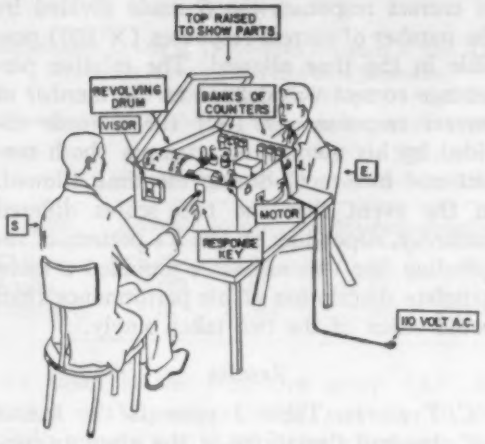

Fig. 1. Semischematic diagram showing the Continuous Performance Test apparatus, a subject and examiner in the positions they would occupy during a test.

Child subgroup ($p < .01$). Also, the IQ of the Non-Brain-Damaged-Adult subgroup was significantly higher than that of the Brain-Damaged-Adult subgroup ($p < .05$). There were no other significant differences between the various paired subgroups with respect to age or IQ. The mean length of institutionalization for the Brain-Damaged-Defective subgroup was 11.4 years, for the Non-Brain-Damaged-Defective subgroup, 9.3 years. This difference was not significant.

Apparatus. The testing instrument is illustrated in Fig. 1. It consisted essentially of a revolving drum on which two series, each of 31 letters, were mounted side by side.

One series served for the *X* task, the other for *AX* task. The drum revolved slowly, approximately twice a minute, in a boxlike case equipped with a visor through which the *S* was required to look. The *E* shifted the task from *X* to *AX* by changing the horizontal position of the visor. The letters were illuminated briefly one at a time. The illumination was provided by a bank of five 0.5-watt neon bulbs mounted inside the case, fired by means of the discharging of an 8-microfarad condenser. The *S* performed by pressing a response key when certain "correct" letters appeared. In the *X* task, every *X* was a correct letter; in the *AX* task, every *X* following an *A* was a correct letter. The letters came into view at approximately 0.92-sec. intervals. A response was scored correct if the response key was pressed within 0.69 seconds after the brief illumination of the correct letter. Following this interval, there was a period of approximately 0.23 seconds before the next letter was illuminated. Correct and incorrect responses were recorded automatically on a set of counters facing the back of the machine where *E* sat. The *X* task was always presented first and included eight *X*'s in the series of thirty-one letters. The responses occurring over the first two revolutions of the drum served as practice and were not counted in the *S*'s score. The *S*s were then given either a 5-minute test (10 revolutions) in which there were 80 possible correct *X* responses, or a 10-minute test (20 revolutions) in which there were 160 possible correct responses. Following a 2-minute rest period, the visor was moved by *E* so that the *AX* series was in view. There were six correct *AX* sequences among the thirty-one letters of the *AX* series. As with the *X* task, the responses occurring during the first two drum revolutions served as practice. Then, the *S*s received either a 5-minute test in which there were 60 possible correct *AX* responses, or a 10-minute test, in which there were 120 possible correct.

Procedure. No attempt was made to keep the instructions completely identical from *S* to *S;* the typical form, however, did not vary greatly from this content:

"Do you see this viewer? When the machine starts, you will see letters appearing one at a time. Your job is to press this key every time you see an *X*. Don't press it for any other letter and always press it when you see an *X*. Press the key down for a bit when you see an *X;* don't hit it too quickly. Try holding your eyes different distances from the viewer until you find the one that is most comfortable for you. Remember, press the key every time you see an *X*, but not for any other letter." Those *S*s who had difficulty in understanding were given further instruction of the same sort.

With the *X* task and the two-minute rest period completed, the *S* was instructed as follows:

"Your job in this next part is again to press the key when you see an *X*, but this time, only if the *X* follows right after an *A*. That is, when you see the series *A—X—* press the key on the *X*. Don't press the key for any other letter following an *A* except *X*, and always press it then. Remember now, when you see the letter *A*, get set; if an *X* comes right after it, press the key." Additional instruction was given to those who had difficulty in understanding.

Except for the children, all *S*s were tested for 10 minutes on each of the two tasks. With children, the time was reduced to 5 minutes on each task, and in addition, the instructions were modified to some extent, presenting the task as a game. For the children under the age of 10, pretesting was carried out to make certain that they could discriminate letters, and to see whether they could follow the in-

structions by requiring answers to two special test sheets, one containing *X*s among other letters, the other, *A*s and *X*s among other letters. Those who failed this pretest were not tested further. No *S* was tested who could not perform the minimal motor tasks of holding his head up unsupported and pressing the response key. Since the testing was conducted at many different clinics and hospitals, it was not feasible to control exactly the amount of illumination in the testing room. The ordinary lighting of the room was used without any modification; care was taken however, to insure that there was no glare on the visor of the testing apparatus.

Reliability. Two measures of the reliability of the CPT were made, using *S*s of the Defective group. Odd-even reliability was determined from the scores of 21 non-brain-damaged *S*s, by comparing the subtotal of correct responses on the 10 even revolutions of the drum with the subtotal on the 10 odd revolutions. The Pearson *r* for the *X* task was 0.88, for the *AX* task 0.86. To determine test-retest reliability, 21 brain-damaged and 22 non-brain-damaged *S*s were retested from four to seven weeks after the initial testing. The test-retest *r* for the number of correct responses of the brain-damaged group on *X* was 0.90, for *AX*, 0.79. The test-retest *r* for the number of correct responses of the non-brain-damaged group on *X* was 0.88, for *AX*, 0.74.

CPT scores. The *X* and *AX* responses of each *S* were scored in two ways. The absolute percentage correct was defined as the number of correct responses the *S* made divided by the number of correct responses (× 100) possible in the time allowed. The relative percentage correct was defined as the number of correct responses (× 100) the *S* made divided by his number of attempts (both correct and incorrect) during the time allowed. In the event that the two scores differed markedly, depending on the *S*'s pattern of responding, the two measures provided a more complete description of his performance than would either of the two taken singly.

Results

CPT scores. Table 2 presents the means and standard deviations of the absolute percentage of possible correct responses on the *X* and *AX* tasks for the six subgroups. Since there was a significant age difference between the Child subgroups and a significant IQ difference between the Adult subgroups (Table 1), the differences between the paired subgroup means on *X* and *AX* in these two groups were evaluated by means of an analysis of covariance. The means of the Child group in Table 2 have been adjusted for regression on age and the means of the Adult group have been adjusted for regression on IQ. The covariance procedure was corrected for heterogeneity of variance between the subgroups when the analysis required it. The mean differences in the Defective group were evaluated by means of the *t* test.

The differences between the subgroup means were highly significant on both the *X* and

Table 2

Adjusted Means and Standard Deviations for Child and Adult Groups, Means and Standard Deviations for Defective Groups, *X* and *AX* Absolute Percentage Correct Measures

	Brain-damaged			Non-brain-damaged			
Group	*N*	Adj. mean	*SD*	*N*	Adj. mean	*SD*	*p* diff.
Children							
X	19	41.61	28.39	26	82.57	14.65	<.0005
AX	19	46.46	28.20	26	79.41	17.44	<.0005
Adults							
X	25	78.53	28.26	25	88.00	17.21	>.30
AX	23	61.40	26.66	23	78.77	15.33	<.01
Defectives							
X	29	46.44	32.02	43	67.86	25.62	<.01
AX	29	32.30	25.79	43	54.74	20.02	<.0005

Table 3

Significance Levels of Differences Between X and AX Scores, for Relative and Absolute Measures of Performance, All Groups

Group	Brain-damaged		Non-brain-damaged	
	Absolute	Relative	Absolute	Relative
Children	>.70	>.30	>.50	>.70
Adults	<.001	<.01	<.01	<.05
Defectives	<.01	<.001	<.001	<.001

Table 4

Significance Levels of Differences Between Adjusted AX Mean Scores of Brain-Damaged and Non-Brain-Damaged Subgroups, All Groups

Group	Performance measure	
	Absolute	Relative
Children	>.20	<.05
Adults	<.01	>.05
Defectives	<.05	<.05

AX tasks in the Defective group ($p < .01$ and $p < .005$, respectively) and also in the Child group ($p < .0005$ in each case). In the Adult group, the brain-damaged subgroup was significantly inferior to its controls on AX ($p < .01$), but not on X ($p > .30$).

When the *relative* percentage scores of the paired Child and Adult subgroups were compared by means of an analysis of covariance, and those of the Defective group by means of the t test, the paired subgroup differences were all in the same direction and of approximately the same magnitude as those reported in Table 2. The levels of significance were in no case less than the corresponding levels reported in Table 2. The only difference was that the *relative* scores on the X task discriminated significantly ($p < .05$) between the Adult subgroups, whereas the absolute scores on the X task (Table 2) did not.

In order to determine if the AX task was in fact more difficult than the X task, the mean X score was compared with the mean AX score for each of the six subgroups. Table 3 presents the significance levels from the t test of the differences between X and AX scores for both the relative and absolute measures of performance in all six subgroups. Each of the Defective and Adult subgroups performed significantly worse (at $p < .05$ or less) on the AX than on the X task. With the children, none of the differences between X and AX was significant. To determine if the brain-damaged subgroups had relatively more difficulty on AX (i.e., independent of their performance on X) than the non-brain-damaged subgroups, an analysis of covariance was run on the AX scores of the paired brain-damaged and non-brain-damaged subgroups with X scores as the covariate. Table 4, which presents the results of the analysis, shows that in terms of the absolute measures

Table 5

Percentages of Correct Identification of Diagnostic Category by CPT Scores, All Groups

Subgroup	Group					
	Defectives		Children		Adults	
	Brain-damaged	Non-brain-damaged	Brain-damaged	Non-brain-damaged	Brain-damaged	Non-brain-damaged
N	29	43	19	26	25	25
$X-R$	75.9***	67.4**	84.2***	76.9***	68.0*	68.0*
$X-A$	62.0	58.1	89.5***	76.9***	64.0	64.0
$AX-R$	72.4**	65.1*	84.2***	76.9***	66.7*	66.7*
$AX-A$	75.9***	67.4**	84.2***	76.9***	65.0†	65.0†

* $p < .05$.
** $p < .01$.
*** $p < .005$.
† Based on 23 subjects.
Note.—p-values refer to the probability that the percentage differs from chance (50%). Nonstarred percentages do not differ significantly from chance.

of performance, the Brain-Damaged-Adult and -Defective subgroups had significantly greater difficulty on the *AX* task than their respective nondamaged subgroups, even when the paired subgroups were adjusted to equal levels of performance on the *X* task. Similarly, in terms of the relative measures of performance, the Brain-Damaged-Child and -Defective subgroups had significantly greater difficulty on *AX* than their respective nondamaged subgroups.

Correct identification of subgroup. Table 5 presents the percentages of correct identification of diagnostic category of all the subgroups on all measures yielded by the CPT. These were determined for each subgroup on each CPT measure from the over-all median of scores of a group on that measure. For example, considering the *X* scores in the Defective group, the percentage of brain-damaged *Ss* scoring *below* the combined Defective group median defines the percentage of correct identification of brain-damaged defectives on the *X* task. Similarly, the percent-

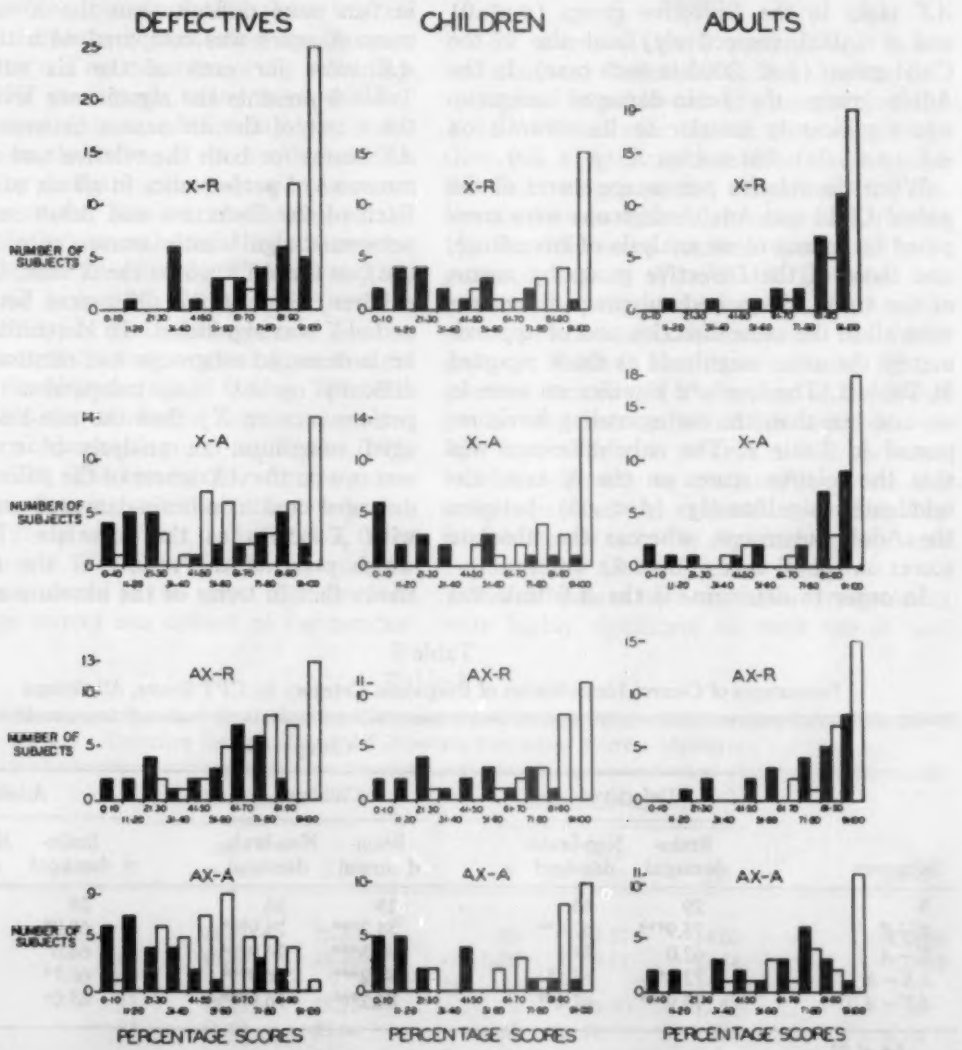

Fig. 2. Frequency distributions of subjects by percentage correct score intervals, all groups, all CPT measures. Black columns represent frequencies of brain-damaged subjects, white columns, frequencies of non-brain-damaged controls.

age of non-brain-damaged defectives scoring *above* the over-all median defines the percentage of correct identification of non-brain-damaged defectives on the *X* task. These percentages were computed to provide an estimate of the potential usefulness of the several scores of the CPT as diagnostic indicators in various clinical groups. In Table 5, *X-R* and *AX-R* refer to the relative measures of performance on these tasks; *X-A* and *AX-A* refer to the absolute measures of performance. The relative measures provided more significant percentages of correct identification than did the absolute measures; all of the six *X-R* and all of the six *AX-R* subgroup percentages were significant at $p < .05$. The Child group had more significant percentages of correct identification than any other (eight of eight) all of which were significant at $p < .005$. There were equal numbers of significant percentages of correct identification among the non-brain-damaged and brain-damaged subgroups; however, these percentages tended to be higher among the brain-damaged subgroups.

Figure 2 presents for each group and CPT measure, the distribution by percentage score of brain-damaged and non-brain-damaged *Ss* (uncorrected for age or IQ differences between subgroups). Figure 2 emphasizes in general the great range in scores achieved by brain-damaged *Ss* as compared with non-brain-damaged *Ss*. The scores of the non-brain-damaged *Ss* are usually concentrated near the upper end of the distribution; in fact, the modal score for non-brain-damaged *Ss* is in the 91 to 100% interval in every case except for the *AX-A* measure in the Defective group. In contrast with this, the scores of the brain-damaged *Ss* span the entire distribution, or the entire distribution excepting one score interval, in every case.

Discussion

The results of this study indicate that brain-damaged individuals perform poorly relative to non-brain-damaged controls on a task requiring continuous attention. The results also indicate that on a task designed to require even more sustained attention the brain-damaged individuals perform even more poorly. The score distributions of the various subgroups (Figure 2) and the percentages of individuals correctly identified by diagnostic category (Table 5) suggest that the CPT is sufficiently sensitive to the effects of brain damage so that it might prove useful clinically. The data of Table 5 suggest that the relative measures are more discriminating than the absolute measures, that the test is somewhat more reliable in reflecting the presence than the absence of brain damage and that it is most discriminating when applied to the children of this study.

These findings are of course preliminary. Use of the CPT in the clinic must await standardization on a variety of diagnostic groups, to ascertain whether brain pathology only and which types of brain pathology lead to deficient performance. For example, some data which have not been reported here indicate that schizophrenics perform no differently from normals of comparable age and IQ. More information concerning other groups is required. Standardization on a variety of diagnostic groups is also necessary in order to have a reliable diagnostic score for each subgroup. Thus, in the present study, cutoff points which included most brain-damaged children missed many brain-damaged adults (see Figure 2).

Diagnostic considerations aside, the question remains as to whether the impaired performance of the brain-damaged *Ss* on the CPT is in fact due to an impairment of attention, whether in the form of momentary lapses, or of a lowering of a general level. Clearly, the available information does not rule out alternate interpretations of the deficit. To make certain that it is attention that is impaired would require more extensive information about the *S*'s test behavior than the present instrument provides. It is clear, furthermore, that the available information does not permit the conclusion that the impaired CPT performance is indeed related to hypersynchrony as evident in the EEG. Nonetheless, the results of this study do lend support to these hypotheses suggesting that it might now be profitable to make a more direct test of the original formulation. This test would be provided by simultaneously measuring CPT performance and brain activity to determine whether errors on the CPT are

coincident in time with hypersynchronous patterns on the EEG and whether there is a correlation between degree of deficit on the CPT and degree of hypersynchrony.

Summary

1. The Continuous Performance Test, a procedure for the detection and study of brain damage in humans, is described.

2. Three groups of *Ss*, each including a brain-damaged and a non-brain-damaged subgroup, were tested on this procedure.

3. The brain-damaged subgroups were significantly inferior to their non-brain-damaged controls on the measures yielded by the CPT, and these differences were increased when the difficulty of the task was increased.

4. The CPT is sufficiently reliable and yields sufficiently large differences between subgroups to suggest that it might ultimately prove useful as a clinical instrument for the diagnosis of brain damage.

5. An interpretation of the inferior performance of the brain-damaged *Ss* was offered in terms of impairment in attention or alertness and suggestions were made about future research relating cerebral events and CPT performance.

Received January 13, 1956.

References

1. Hebb, D. O. *The organization of behavior.* New York: Wiley, 1949.
2. Levin, S., Greenblatt, M., Healey, M. H., & Solomon, H. C. Electroencephalographic and clinical effect of prefrontal lobotomy, with consideration of postlobotomy convulsive seizures. In M. Greenblatt, R. Arnot, & H. C. Solomon (Eds.), *Studies in lobotomy.* New York: Grune & Stratton, 1950. Pp. 400–427.
3. Lindsley, D. B. Electroencephalography. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* New York: Ronald, 1944. Pp. 1033–1103.
4. Mettler, F. A. (Ed.) *Psychosurgical problems.* New York: Blakiston, 1952.
5. Mettler, F. A. (Ed.) *Selective partial ablation of the frontal cortex.* New York: Hoeber, 1949.
6. Partridge, M. *Pre-frontal leucotomy.* Springfield, Ill.: Charles C Thomas, 1950.
7. Petrie, A. *Personality and the frontal lobes.* New York: Blakiston, 1952.
8. Schwab, R. S. *Electroencephalography in clinical practice.* Philadelphia: Saunders, 1951.

Journal of Consulting Psychology
Vol. 20, No. 5, 1956

Stimulus Structure as a Determinant of the Perceptual Response[1]

M. Powell Lawton
Veterans Administration Hospital, Providence, R. I.

The role of formal stimulus characteristics is acknowledged to be of importance in the interpretation of projective test data. Clarification of the meaning of projective test responses has been sought through study of stimulus qualities in a number of areas (1, 2, 6, 8, 12, 14, 15).

With respect to the Rorschach test, the role of form in the production of a response has received surprisingly little attention. The concept of the popular response implies that formal blot characteristics of certain areas of the Rorschach tend to call forth a particular response. Ranzoni, Grant, and Ives (10) obtained a set of norms for an adolescent population specifying which cards were most likely to elicit responses using a particular location, determinant, and content. Their study was not concerned primarily with form, however, and made no attempt to vary the forms of the Rorschach stimuli or to place them in an order of complexity. Meer's study (9) established a tentative order of organizational difficulty among cards, but did not attempt to specify the factors believed to be responsible for stimulus ease or difficulty. Stein's (11) tachistoscopic exposure of the cards demonstrated that perception was less adequate under impoverished stimulus conditions but again, the stimulus condition as a whole was varied, with no attempt to focus on form as the variable for study.

The present study investigated by means of a non-Rorschach perceptual task the general hypothesis that stimulus ambiguity is an important determinant of perceptual adequacy. Degree of maladjustment was a second variable studied in line with Forer's (5) suggestion that degree of personality integration is a determining factor in the differential adequacy with which an individual perceives in situations of varying ambiguity.

It was hoped that variation of the form of the stimulus, isolated from as many other stimulus characteristics as possible, would provide a general basis upon which to interpret individual projective records, though the limits of applicability to any specific test are fully realized. As in Stein's (11) study, response characteristics were evaluated in terms of perceptual adequacy, defined so as to include both formal and content characteristics of the response. The specific hypothesis was that both increasing stimulus ambiguity and increasing degree of pathology would result in decreased perceptual adequacy.

Procedure

In order to limit investigation as much as possible to form alone, sixty figures were randomly cut from black paper and mounted on white cards, leaving only the blackness of the figure as a confounding factor. Seventeen psychologists rated each figure as most, least, or moderately structured, using as their criterion of structuredness the relative ease with which each figure suggested a percept. From these judgments five figures were clearly chosen as

[1] Adapted from a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Faculty of Philosophy, Columbia University, 1952. The author wishes to thank Dr. E. J. Shoben, Jr., Dr. Helen Walker, and Dr. Laurance Shaffer, members of his committee, for their help, and also Dr. Bernard Locke, Chief Psychologist, New York VA Hospital, for making possible the gathering of the data.

the most structured (Structure I), five as moderately structured (Structure II), and five as least structured (Structure III). The 15 figures were arranged in random order for standardized presentation to the subjects. This "subjective consensus" method of defining stimulus ambiguity, though it has the disadvantage of risking bias through the use of only psychologists as judges, affords a definition of ambiguity in psychological, rather than physical, terms. The psychological definition was deemed more appropriate to the present study.

Thirty *Ss* were chosen from each of three nosological groups: neurotic, latent schizophrenic, and hospitalized schizophrenic, as diagnosed either by their outpatient therapists or ward physicians. By using this ordinal scale of degree of maladjustment, it was hoped that unreliability in ratings of maladjustment could be avoided. The three groups were matched for age and education. Analysis of variance and Bartlett's test showed no reason to suspect that either means or variances differed significantly. The mean age of the entire group was 29.8 years, mean educational level 11.4 years.

The cards were presented in standard position and *Ss* were instructed to tell the examiner what the whole figure could look like. Turning and detail responses were discouraged, though accepted if *S* could not respond in the requested manner. Only the first response to each figure was used. Through judicious urging only five rejections out of a possible 1,350 responses occurred. The *S* was asked to describe each response further. This was the examiner's only question, other than clarification of the areas used.

Table 1

Number of Patients Receiving All "*1*" Ratings on Form Level

Group	Stimulus Ambiguity I	II	III
Neurotics	28	19	14
Latent schizophrenics	25	13	8
Hospitalized schizophrenics	17	3	1

The following categories were employed in evaluating the perceptual quality of the responses:

Form Level: a four-point rating scale measuring judged goodness of fit of percept to stimulus. Scale points were named *1.* adequate form, *2.* vague form, *3.* poor form, and *4.* very poor form.

Structural Peculiarities: (*a*) *Fabulation* was a response which added a great deal more to the percept than could reasonably be assumed from the stimulus ("Miss Olsen's profile," "a person who died of pneumonia.") (*b*) *Incongruity* was a percept whose parts do not belong together ("a person with a meat cleaver attached to his head," "a rabbit—here's the opening in his shell.")

Sex Response: Any overt genital or anal response.

Other content categories were scored, but occurred infrequently and could not be scored reliably. A Total Pathology rating scale was also used, but because of its high correlation with the Form Level scale, added little further information.

Each response was typed and coded individually and the ninety responses to each figure were scored consecutively by the writer. One-third of the total responses were submitted for a reliability check to two other

Table 2

Number of Patients Without Structural Peculiarities, With Cochran's *Q* Test and Combined Probability Measure

Group	Stimulus ambiguity I	II	III	*Q*	*p*	Order probability	Combined probability
Neurotics	29	28	26	2.36	.308	1/6	.114
Latent schizophrenics	21	19	15	3.10	.214	1/6	.080
Hospitalized schizophrenics	20	19	18	.48	.794	1/6	.295
All groups	70	66	59	5.03	.082	1/6	.031

psychologists.[2] Through a method described in detail elsewhere (7), scoring reliability was determined and was deemed good enough to permit using the writer's scoring in the evaluation of the data.

Results

Table 1 shows the number of patients all of whose responses received Form Level scores of *"1,"* by clinical group and by stimulus structure. The number of patients receiving all *"1"* ratings decreased regularly both as stimuli became more ambiguous and as pathology increased. A nonparametric significance test was performed as follows. For nine cells there are 9! possible arrangements, and of these 9! arrangements only 42 can be found such that in every row the frequencies decrease from left to right and in every column decrease from top to bottom. Accordingly, the probability of obtaining by chance alone data showing this regular decrease in Form Level with increasingly ambiguity and with increasing degree of illness is 42/9!, or .0001156.

Table 1 suggested the possibility that the rate of decline in Form Level as stimuli become more ambiguous might be greater as psychopathology increases. This supposition was tested in the following manner. A rate of decline score was obtained for each *S* by subtracting the sum of his Form Level scores for Structure I from the sum of his Form Level scores for Structure III and dividing this difference by the sum of Structure I and Structure III Form Level sums. The mean rate of decline score for neurotics was .0759, for latent schizophrenics .1223, and for hospitalized schizophrenics .1935. An analysis of variance resulted in an *F* value of 4.95, significant at the .01 level for 2 and 87 degrees of freedom.

[2] Dr. Gerald Bauman and Dr. Melvin Kornreich most kindly acted as judges.

Table 3

Number of Patients Who Give No Structural Peculiarity

Group	Number*
Neurotics	25
Latent schizophrenics	11
Hospitalized schizophrenics	12

* Chi square = 7.61; $p < .05$.

The incidence of Structural Peculiarities is shown in Table 2. Columns I, II, and III indicate the number of patients who did not respond with a Structural Peculiarity to any one of the five figures of each structural level. In order to test the hypothesis that incidence of Structural Peculiarities is a function of stimulus ambiguity, Cochran's *Q* test (3), a variant of chi square that does not require independence of observations, was used. None of the probabilities (Column six, Table 2) associated with *Q* values reaches statistical significance, but when account is taken of the fact that the predicted order across rows was obtained in each case (13), the probability value for the entire sample is seen to be .031. The usual chi-square test applied to the relationship between degree of maladjustment and incidence of Structural Peculiarities (Table 3) gave a chi-square value of 7.61, associated with a probability value less than .05. Inspection reveals that the significance of this measure is derived largely from the relative absence of Structural Peculiarities in the neurotic group.

Table 4 indicates the number of patients showing no Sex responses to cards of different stimulus structures. Again using Cochran's

Table 4

Number of Patients Without Sex Responses, With Cochran's *Q* Test and Combined Probability Measure

	Stimulus ambiguity						
	I	II	III	*Q*	*p*	Order probability	Combined probability
All groups	86	84	81	3.80	.142	1/6	.053

Table 5

Number of Patients Who Give No Sex Response

Group	Number*
Neurotics	29
Latent schizophrenics	22
Hospitalized schizophrenics	25

* Chi square = .98; $p > .05$.

Q test and combining probabilities, the association between stimulus structure and incidence of Sex responses is seen to be of borderline significance ($p = .053$).

Table 5 indicates the number of patients giving no Sex responses. The chi-square value of .98 is associated with a probability greater than .05. Accordingly the hypothesis that incidence of Sex responses is a function of degree of maladjustment must be rejected.

Discussion

Perceptual distortion as measured by form accuracy is thus seen to be a function of both stimulus ambiguity and degree of maladjustment. On the other hand, perceptual distortion as measured by indices relating to content and the organization of content show a highly tenuous, if at all significant, relationship to stimulus ambiguity and to degree of maladjustment.

It appears clear, then, that it is important to consider stimulus characteristics in the interpretation of a projective response. In the Rorschach, for example, a response judged to be of poor form quality would have different meaning depending upon whether it occurred to a highly structured area or to a relatively amorphous area. As indicated in Table 1, a poor form response to the most ambiguous stimuli would give less information about the presence of a schizophrenic process in an individual than would a poor form response to a highly structured stimulus: the original data show that four neurotics gave ten definitely poor form responses to the less well-structured cards, but none to the best structured. It can be inferred that any response of less than adequate form level to a popular area of the Rorschach cards should make the interpreter suspicious of a severe disturbance, whereas an accumulation of $F-$ responses on less well-structured areas would be required to give the same impression. The present results would tend to confirm Meer's (9) supposition that an important factor in determining Rorschach card difficulty is stimulus ambiguity. A problem for further research might involve the development of a more discriminating Rorschach $F+\%$ than is presently used by weighting more heavily those $F-$ responses that occur to highly structured areas.

On the other hand, the equivocal results obtained in the attempt to relate Structural Peculiarities and Sex responses to stimulus structure make interpretation somewhat difficult. It may be that there is actually a greater contribution of internal factors than of stimulus factors in the production of these content-related distortions, though the present study does not answer this question. This supposition is reasonable if one considers that standards of what is appropriate in matters of content are largely internal, in distinction to the Form Level dimension, where the model against which the appropriateness of the *S*'s association is tested is physically present in the stimulus.

The differential rates of decline in Form Level with increasing ambiguity among the three clinical groups offers experimental verification of the observation that the more maladjusted individuals "go to pieces" more easily than the less maladjusted. Not only is there a difference in adaptive ability, but this difference becomes wider as the external situation becomes more demanding, in the sense of forcing the individual to provide structure to an ambiguous field from within himself. Such a finding would suggest that the best differentiation among groups according to degree of maladjustment would come from rating the Form Level of the more ambiguous cards. This supposition was tested by Feldman, Gursslin, Kaplan, and Sharlock (4) who found that an $F+\%$ which excluded popular responses from consideration was more discriminating between normals and neurotics than Beck's or Hertz's $F+\%$.

Summary

As a means of exploring the Form Level dimension in projective perception, a method

was devised to test the hypothesis that both increasing stimulus ambiguity and increasing personality maladjustment would have an adverse effect upon perceptual adequacy. Black paper cutouts of three ranked degrees of ambiguity were presented in Rorschach fashion to three groups of patients representing three degrees of maladjustment. It was found that both greater stimulus ambiguity and greater personality maladjustment were associated with lesser perceptual adequacy, as measured by Form Level. Perceptual adequacy as measured by several indices involving response content and organization of content was seen to have an equivocal relationship to stimulus ambiguity and to personality maladjustment. It was concluded that stimulus characteristics are often important determinants of projective responses and that consideration of form ambiguity is necessary to the best interpretation of projective records.

Received November 21, 1955.

References

1. Allen, R. M. The influence of color in the Rorschach test on reaction time in a normal population. *J. proj. Tech.*, 1951, **15**, 481–485.
2. Bijou, S., & Kenney, D. T. The ambiguity values of TAT cards. *J. consult. Psychol.*, 1951, **15**, 203–209.
3. Cochran, W. The comparison of percentages in matched samples. *Biometrika*, 1950, **37**, 256–266.
4. Feldman, M. J., Gurssiin, C., Kaplan, M. L., & Sharlock, N. A preliminary study to develop a more discriminating $F+$ ratio. *J. clin. Psychol.*, 1954, **10**, 47–51.
5. Forer, B. The latency of latent schizophrenia. *J. proj. Tech.*, 1950, **14**, 297–302.
6. Kenney, D. T., & Bijou, S. Ambiguity of pictures and extent of personality factors in fantasy responses. *J. consult. Psychol.*, 1953, **17**, 283–288.
7. Lawton, M. P. Stimulus ambiguity and personality maladjustment as related to perceptual adequacy. Unpublished doctor's dissertation. Columbia Univer., 1952.
8. Lazarus, R. The influence of color on the protocol of the Rorschach test. *J. abnorm. soc. Psychol.*, 1949, **44**, 506–516.
9. Meer, B. The relative difficulty of the Rorschach cards. *J. proj. Tech.*, 1955, **19**, 43–53.
10. Ranzoni, J., Grant, M., & Ives, V. Rorschach "card-pull" in a normal adolescent group. *J. proj. Tech.*, 1950, **14**, 107–133.
11. Stein, M. Personality factors involved in the temporal development of Rorschach responses. *J. proj. Tech.*, 1949, **13**, 353–414.
12. Wallen, R. The nature of color shock. *J. abnorm. soc. Psychol.*, 1948, **43**, 346–356.
13. Wallis, W. Compounding probabilities from independent significance tests. *Econometrika*, 1942, **10**, 229–248.
14. Weisskopf, Edith. A transcendence index as a proposed measure in the TAT. *J. Psychol.*, 1950, **29**, 379–390.
15. Weisskopf, Edith. An experimental study of the effect of brightness and ambiguity on projection in the TAT. *J. Psychol.*, 1950, **29**, 407–416.

Journal of Consulting Psychology
Vol. 20, No. 5, 1956

Manifest Anxiety and the Palmar Perspiration Index[1]

Allen D. Calvin, F. J. McGuigan, Sybil Tyrrell, and Marshall Soyars
Hollins College

During the last few years the Taylor A scale has become an increasingly important measure of manifest anxiety. During the same period Mowrer and others from the University of Illinois have been developing the Palmar Perspiration Index (PPI) as a measure of psychological tension. The present study is an attempt to investigate the relationship between the A scale and the PPI.

The A scale and the higher form of the Otis were given to 54 female undergraduate students. In order to create a stress situation the *Ss* were called in for individual experimental sessions and given the Wechsler digit span. Immediately after completing the digit span a PPI was obtained. After approximately three months *Ss* were asked to fill out a six-point self-rating anxiety scale.

Product-moment correlation coefficients were obtained between the following: A scale and PPI $-.01$; A scale and digit span $-.05$; digit span and PPI $-.16$; Otis and A scale $-.02$; Otis and PPI .22. None of these was significant. A biserial correlation of .35 ($p < .05$) was obtained between self-ratings and A scale scores, and a biserial of .17 (not significant) was obtained between self-ratings and the PPI.

The failure to find a correlation between A-scale scores and PPI scores was unexpected since at first glance both seem to be measuring somewhat the same variable. However Beam (1) in a study published after the completion of our experiment, reports that he too found no significant relationship between the PPI and A scale. He points out that his Taylor scores fell within the middle range and were fairly homogeneous. Our A-scale scores had a mean of 16.9 with an *SD* of 6.9, indicating considerable scatter within our group. It seems highly unlikely that our failure to obtain a significant relationship between A scale and PPI scores can be ascribed to too much homogeneity within our group.

On the basis of our findings and other studies of anxiety the following hypothesis is proposed: The A scale and the PPI measure two different facets of behavior. Where anxiety is of a temporary "ego-involved" situational nature, the appropriate measure of drive is the PPI. Conversely, where anxiety is a stable permanent personality characteristic of the individual the appropriate measure of drive is the A scale.

Brief Report.
Received April 26, 1956.

[1] An extended report of this study may be obtained without charge from Allen D. Calvin, Hollins College, Va., or for a fee from the American Documentation Institute. Order Document No. 4849 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

Reference

1. Beam, J. C. Serial learning and conditioning under real-life stress. *J. abnorm. soc. Psychol.*, 1955, **51**, 543–551.

Journal of Consulting Psychology
Vol. 20, No. 5, 1956

Effective Intelligence as Measured by an Unstructured Sentence-Completion Technique[1]

George E. Copple
Vanderbilt University

Numerous varieties of test material and test procedure have been used in assessing the intellectual functioning of children. Almost all these procedures have involved rather strict structuring, and this structuring has ordinarily been considered a merit of the procedure. Perhaps because of this emphasis on easily scored, "right or wrong" answers, the possibilities of using a sentence-completion technique to evaluate intelligence have been largely overlooked. The present study will describe a sentence-completion technique which has been developed as a measure of "effective intelligence" and will present data on its usefulness.

Earlier Uses of Sentence Completions

Beginning with Ebbinghaus (1), and best known from its inclusion on the Stanford-Binet (2), one type of sentence completion has long been used as an indicator of intelligence. In this type, most of the sentence is given, with only a connective word omitted. Acceptable choices for this omitted word are quite limited and an answer is either "right" or "wrong." M. R. Trabue (3) used a similar method in a large-scale study. He called his test "language scales" and it is obvious that knowledge of word meanings and language usage played the greatest part in determining a successful performance.

The other, and more recent, use of sentence completions has been to evaluate personality or adjustment. Every survey of projective techniques is likely to have a section or a chapter on sentence-completion techniques. These techniques are usually for adults, they consist of "loaded" stems intended to tap specific areas of adjustment ("My greatest weakness is . . ." or "I think that most mothers . . ."), and the responses are evaluated for content, particularly dynamic content. The level of ability shown by the completions and the range and depth of interests are ignored.

The use of the sentence-completion technique to be described in the present paper is different from these other uses reported in the literature in general purpose, technique of administration, and rationale.

Rationale

Most intelligence-test items demand little that is creative or even integrative of the person being tested. They provide a task or a problem that has a narrow range of correct answers. They provide a great deal of structuring and often call for a single word or phrase or brief explanation. Typical items are: "What is the thing to do if you are on your way to school and see that you are in danger of being late?" "What is the color of rubies?" "How are a coat and a dress alike or the same?" "*Sky* is to *blue* as *grass* is to ———." Undoubtedly items of this kind give some valid indication of a subject's intelligence. But when a test is composed exclusively or mostly of such items, some important aspects of intelligence are left out. The individual who can readily place Brazil in South America, Egypt in Africa, or India in Asia, may not be able to respond successfully if asked instead to name one country in each of these continents. With still less structur-

[1] This article represents a revision of a paper presented at a meeting of the Southern Society for Philosophy and Psychology in Atlanta, Georgia, on April 16, 1954.

ing, as for example in responding to the instruction to give the best sentence completion he can to the stimulus "In other countries . . .," he may give us something inept and vague such as "things are different" or "they are foreigners." One limitation of virtually all intelligence tests is just this failure to let the subject do enough—the failure to let him show us what he can do and does do when the stage is not too completely set.

An example may be useful. Here are two nine-year-old boys who are completing the stimulus-stem "A piece of string. . . ." One completes it "is long." The other supplies "can be used to make a bow and arrow." This latter is obviously a far superior answer. It is most unlikely, however, that the first boy was ignorant of the fact which the second boy's completion utilized. On a more conventional test item calling for this information, such as, "A bow is made of ——— and ———," both boys would probably be able to answer correctly. But does not the second boy's method of completing the sentence show something in addition to the possession of information? Does not this "something" seem to be basic to many types of school and nonschool performance that we label "intelligence"?

Administration of the Test

The Copple Sentence Completion Test is composed of 60 items, only 30 of which are given to any one subject. Level A, consisting of items 1–30, is for children 5–8 inclusive. Level B, consisting of items 16–45, is for children 9–12 inclusive. Level C, with items 31–60, is for children 13 and older, and for adults. Items at the various levels differ somewhat in difficulty, but this is not very easy to insure and is of minor importance. The scoring standards differ for the different levels of the test, so that even if an intrinsically easy item appears on the higher level test, it is more strictly scored and a better performance is required for a comparable rating.

Administration is always individual, with oral responses by the subject, which are recorded by the examiner. The procedure enables young children or illiterates to respond as fully as they are able, without penalizing or limiting them in any way. It permits people in all age groups to use natural conversational completions without regard to avoiding words they cannot spell. At the other extreme, it discourages "literary" completions. Experience seems to show that it is unnecessary to limit the time a subject can use in framing his responses. As a rule, the longer the delay, the poorer the response (except for extremely brief latencies).

In introducing the test, the subject is told that he will be given some words and that he is to add some more words so as to produce the very best sentence he can make. A sample stem is given the subject and two ways of completing it are presented, one being designated as "better" than the other. For example, the stem might be "The house. . . ." One way to complete it, the subject is told, would be to say "is large." Another, and a better way, would be to say "stood in the bend of the road." Nothing is said about *why* the latter is better, and subjects as young as five years accept this as a self-evident superiority.

With this brief introduction, the test is begun. It requires about 15 minutes to complete. It is rare, even at age five, to have a child say "I don't know" to even a single item. Few persons seem to experience any sense of failure in completing the sentences. The intrinsic interest is high and no matter at what level the individual is performing, he is likely to consider that he is doing well. This feature, of course, is a particular virtue when a whole battery of tests is to be given or when a subject is to be worked with therapeutically.

Scoring the Responses

Each response is scored 0, 1, 2, or 3 in accordance with prepared scoring standards. These standards reflect such considerations as relevance, complexity, descriptiveness, and range or scope. Numerous examples have been provided for each score value on each item. Judgment is still necessary, but in numerous studies the agreement in scoring has been high ($r = .95$ or above). For example, in one study using 24 protocols and with 10 psychology graduate students as scorers, none of whom had had any previous experience with the test, the median interscorer reliability was .97.

The split-half reliabilities of the test itself have been uniformly high. For the three studies reported below, these reliabilities (corrected by the Spearman-Brown formula) were respectively .94, .95, and .93. No test-retest data have been gathered.

Measuring "Effective Intelligence" (Validity)

The most valuable criterion for a test of "effective intelligence" would be the judg-

Table 1

CSCT Scores for Three Age Groups

Age Group	*N*	Mean Test Score	Sigma	*CR*
9–6	29	40.76	8.98	2.27 (.02 level)
10–6	17	47.71	10.14	1.66 (.05 level)
11–6	20	53.85	11.73	

ments of objective, well-trained observers on the actual behavioral efficiency of large numbers of individuals carrying out a variety of tasks. Such a criterion is not available, and the "construct validity" of a measure of effective intelligence must be examined by diverse approaches.

Relationship of test scores to age. Two studies with data relevant to the relationship between test scores and CA can be cited. In the first, Form B of the CSCT was administered to all the children in two representative schools in Davidson County, Tennessee, a metropolitan community, who were within three weeks of being 9–6, 10–6, and 11–6 years of age. Table 1 contains the data for the 66 children in this study. As may be seen, the two older groups of the study are rather small. If they are combined and compared with the younger group, the difference in mean scores is significant well beyond the .001 level.

A second study of the relationship of CSCT scores to age was carried out with a group of 50 orphanage children, again using Form B of the CSCT. These children ranged in CA months from 97 to 154, with a mean of 129.66 and a sigma of 15.7. The obtained correlation between chronological age and CSCT scores is significant at the .0001 level, but this is not very informative. The size of an *r* between chronological age and any measure of mental age will be greatly influenced by the range and distribution of chronological ages included. What is really needed, therefore, is an indication of the *r*s between CAs and mental ages on other well-known tests within this same group. This is available on this group of 50 children for the Goodenough Draw-A-Man Test, but for the Binet, only 33 of the 50 subjects were tested.

Table 2

Correlations Between Age and Scores on Three Different Tests

Test	Sample A ($N=33$)	Sample A′ (A plus 17 additional)
CSCT (Form B)	.584	.577
Binet (Form L)	.621	
Goodenough DAM	.294	.400

All of the tests—Form L of the Binet, the Goodenough, and the CSCT—were administered within a period of five or six weeks. The Goodenough drawings were collected and scored by an advanced graduate student, and the same student administered the CSCT, which was, however, scored by the author. In Table 2 the correlations of each of the tests with ages in months are shown. In computing these correlations, mental ages were used for the Binet and Goodenough data.

Relationship of test scores to scholastic achievement. A widely used criterion of an intelligence test's validity is the correlation of the test with school achievement. Form B of the CSCT and Forms L and M of the Binet were given, plus the Stanford Achievement Test, to a group of 67 orphanage children aged 9–13, none of whom were in the study described above. As a group these children were retarded, as shown by a Binet IQ range (L and M were averaged) of 67.5 to 98. Table 3 summarizes the pertinent data on this group. The range of intellectual ability and of scholastic attainment within this group was rather narrow, considerably less than would be found within a single grade in the typical school system. In spite of this, both the Binet and the CSCT had substan-

Table 3

Range and Variability of an Orphanage Group on Two Tests of Intelligence and an Achievement Test

Test	Range	Mean	Sigma
Binet (L and M) MA	140–76	107.36	13.87
CSCT (Form B)	69–10	39.86	10.81
Stanford Achievement (grade equivalent)	5.4–1.5	3.22	.87

tial correlations with the Stanford scores. The Binet correlation was .623 and the CSCT correlation was .542. Since the Binet MAs and the CSCT were not highly correlated with each other (.349), a multiple *r* between these two and the Stanford Achievement was computed and was found to be .713. While the Binet was slightly more successful than the CSCT in predicting Stanford Achievement scores, it should be noted that the Binet scores involved were averages of two forms, requiring a minimum testing time of two hours for each child, while the CSCT required about fifteen minutes for each child.

Relationship of test scores to IQ gains in therapy. One other set of correlations may have particular interest for clinicians. As part of a larger experiment designed to increase IQs, 55 of these orphanage children with below-average IQs were exposed to an 18-month program of group play therapy, with 4 children in each group, and of enriched cultural opportunities. Both L and M of the Binet were given prior to and again at the conclusion of therapy. The range of IQ gains was not spectacular, from -3.5 to $+17$, with a mean of $+8.26$ and an *SD* of 4.15. There was no relationship (.014) between the original IQ level and the amount of gain. However, the original MA level correlated .172 with the IQ gains and the CSCT scores correlated .295 with these gains. These *r*s are significant at the .11 and the .02 levels, respectively. If other corroboration of this finding is forthcoming, it is potentially important, for any brief procedure which can help to predict the gains of retarded children from therapy is worth knowing about.

Summary

A method of testing "effective intelligence" by an unstructured oral sentence-completion technique (Copple Sentence Completion Test) has been described. A median interscorer reliability of .97 was found in an investigation involving 10 scorers, and in other, smaller projects, the results were comparable. Split-half reliabilities for three studies reported in the present paper were .94, .95, and .93.

Three approaches to the test's validity were reported:

1. CSCT scores compared favorably with Binet and Goodenough mental ages in their correlation with chronological age.
2. CSCT scores compared favorably with Binet mental ages in predicting Stanford Achievement Test scores.
3. CSCT scores were significantly related to IQ gains subsequently made by a group of retarded children undergoing play therapy.

Received November 1, 1955.

References

1. Ebbinghaus, H. Ueber eine neue Methode zue Prufung Geistiger Fahigkeieten und ihre Anwendung bei Schulkindern. *Z. Psychol. Physiol. Sinnesorgane,* 1897, 13, 401–459.
2. Terman, L. M., & Merrill, Maud A. *Measuring intelligence.* Boston: Houghton Mifflin, 1937.
3. Trabue, M. R. Completion-test language scales. *Teach. Coll. Contr. Educ.,* 1916, No. 77.

Journal of Consulting Psychology
Vol. 20, No. 5, 1956

A Note on Porteus Maze and Wechsler-Bellevue Scores as Related to Antisocial Behavior[1]

Kenneth Purcell [2]
Mental Hygiene Consultation Service, Fort Dix, New Jersey

As part of a study on antisocial behavior and some of its personality correlates, data were collected on Porteus Maze and short-form Wechsler-Bellevue performance. Previous studies (1, 2, 6, 9), using the qualitative method for scoring the Maze Test described by Porteus (5), have consistently reflected differences between delinquents and nondelinquents. Nevertheless, it seemed worth while to include the Maze Test in the present investigation, mainly because the type of population used differed in important respects from earlier populations studied. Each of the former investigations had compared institutionalized delinquents with noninstitutionalized, presumably normal nondelinquent youths. The population employed in this study consisted of noninstitutionalized young male adults, all of whom were referred to a mental hygiene clinic for psychiatric problems. Porteus (7) has cited evidence to show that the Maze Test successfully predicts general social adjustment and adaptation. The question posed in this paper concerns the capacity of the Maze Test, qualitatively scored, to distinguish poor social adjustment characterized primarily by antisocial behavior from other types of poor social adjustment.

Hunt and French (3), in one of their reports on the CVS abbreviated intelligence scale, found a tendency for psychopaths and disciplinary cases to score lower on Comprehension than on Similarities and Vocabulary. However, the differences found were not statistically significant. It is suggested by the authors that this may well have been because of inadequate criteria for defining a psychopathic population. Since the groups used in this study were reliably separated along the dimension of antisocial behavior, a further examination of CVS results seemed appropriate. In addition, on the basis of our experience at this clinic, it was felt that the Block Design subtest generally offered the most valid estimate of intellectual capacity among antisocial subjects. This appeared to be because negativism and resistance were minimal and interest maximal for this subtest which Wechsler (8) reports as having one of the highest correlations with Full Scale IQ of all subtests (.73).

Method

A total of 56 male Army trainees between the ages of 17 and 26 with a mean age of 20 were used as *Ss* for this experiment. All were psychiatric referrals to the mental hygiene clinic for whom psychological testing was requested. The only patients excluded were those suffering from organic brain damage or those diagnosed as psychotic.

It was the intent of this study to relate psychological test results to overt behavior. Therefore, each *S* was rated on manifest antisocial behavior trends on the basis of social history data routinely obtained by one of the psychiatric social work specialists. Both military and civilian data were used in making this evaluation. Originally two raters independently judged the degree of antisocial behavior on a seven-point scale with a reliability of .91 being obtained. This scale was later consolidated into a three-point scale to

[1] The author wishes to express his appreciation to Carl Eisdorfer, Marvin Greenstein, Maurice Kouguell, and Dr. Philip Kubzansky who collected data for this study and assisted in its statistical evaluation.

[2] Now at the University of Kentucky.

Table 1

Analysis of Variance of Qualitative Maze Scores of Three Groups Differing in Antisocial Behavior

Group	Mean*		Variance†	*F*	*p*
Group I	30.38	Between groups	3,543.97	3.39	.05
Group II	42.94	Within groups	1,044.15		
Group III	58.57				

* *Ns* = 24, 18, and 14 for Groups I, II, and III, respectively.
† Between groups *df* = 2; Within groups *df* = 54.

secure an adequate number of *Ss* for each category. Group I, the non-antisocial group, was composed of *Ss* whose overt aggressive behavior was minimal with no record of military or civilian disciplinary infractions. Group II, the intermediate group, consisted of *Ss* who may have been in fights with moderate frequency and whose record was marked by some school truancy, perhaps minor infractions of the law, and AWOL from the military. Group III, the antisocial group, consisted of *Ss* with an extensive pattern of delinquency including some criminal actions such as assault, rape, burglary, dope peddling, etc.

The three groups were compared on qualitative Maze Test scores and on the four Wechsler-Bellevue subtests noted above. In each case these measures were administered as part of a battery of clinical tests.

Results and Discussion

Analysis of variance of the qualitative maze scores, as shown in Table 1, discloses a difference between the mean *Q* scores (total weighted qualitative score) of the three groups significant at the .05 level. This is not so significant a result as has been reported in other studies of the Maze Test. For example, Docter and Winder (1) report a difference between delinquents and nondelinquents significant at the .0001 level. There are at least two reasons for the apparent insensitivity of the maze in differentiating between these two groups in the present study. First, the range of *Ss* was marked by more continuity with respect to antisocial behavior than was true in other investigations. The presence of a group intermediate between the extreme antisocial and non-antisocial groups created additional overlap of scores. Second, the number of *Ss* used in this study, particularly if one considers only the two extreme groups, was considerably smaller than the *Ns* reported by other authors (1, 2, 5, 6, 9). If similar differences were obtained with larger *Ns*, the significance level would, of course, be higher.

It will be remembered that Group I *Ss* were also psychiatric referrals, the great majority of whom had manifested some form of social maladjustment in the Army situation. They may have been seclusive, overdependent, easily frightened by authority or socially deficient in almost any respect except that none of these *Ss* were aggressively antisocial in their overt behavior. Since the *Q* scores of these *Ss* were only slightly higher, and on a relative basis not at all higher than the *Q* scores of normal controls used in previous studies, it is apparent that the qualitative Maze Test tends to be sensitive only to certain types of social maladjustment. It seems to be maladjustment marked by nonconforming impulsivity that is singled out by this qualitative analysis. Impulsivity is probably the essential ingredient for high *Q* score. For example, one of the severely antisocial *Ss* was a shrewd, calculating relatively controlled person whose record reflected many of these characteristics. He had been arrested on numerous occasions but always managed to elude conviction. His *Q* score of 5 was one of the lowest for all *Ss*. In his behavioral notes the tester noted that *S* was calm, detached, and obviously out to demonstrate that he (*S*) could easily master the situation.

Since several authors (1, 7) have noted a small but significant correlation between *Q* score and intelligence, the intellectual factor must be measured and accounted for in some way. The task of estimating the intellectual

capacity of antisocial *Ss* is difficult since their test performance is so colored by uncooperative, negative attitudes. This was a particularly important problem in this study, as many of the *Ss* were actively engaged in intense rebellion against any form of military authority.

Table 2 presents the results of the analysis of variance of the Wechsler-Bellevue data. The differences between the means of the three groups are significant for Prorated IQ, Similarities, Vocabulary, and Comprehension scores with antisocial *Ss* achieving the lowest scores in each instance. By contrast, the differences found in Block Design means do not even approach significance. The reasons for this finding are at least twofold. First, almost every antisocial *S* had been an habitual truant and had left school much earlier than the non-antisocial *Ss*. The benefits of educational experience on such tests as Vocabulary or Comprehension were lost to these *Ss*. Second, the verbal interaction required by the verbal subtests reproduces former unpleasant school situations and creates many more opportunities for the arousal of authority conflicts than occurs during administration of the Block Design test. During the latter, *S* may focus his attention on impersonal objects without the necessity of replying directly to the examiner. Antisocial individuals seem to find it less necessary to assert their rebellious nonconformity in this more oblique relationship with the examiner.

Turning briefly to a comparison of Comprehension scores with those of Similarities and Vocabulary, we find, from an inspection of means and significance levels indicated in Table 2, the same tendency for antisocial *Ss* to score relatively lower on Comprehension, as was noted by Hunt and French (3). Although the differences between these subtests do not attain statistical significance, the qualitative variation between the groups in their responses to certain items of the Comprehension test seems worthy of note. The answers of antisocial *Ss* to the questions on the envelope, bad company, taxes, and law, commonly contained antisocial content in some part of the response, although giving credit to the response might be required under the rules of scoring. Perhaps a revised scoring scheme or the remodeling of Comprehension to encourage antisocial answers in

Table 2

Analysis of Variance of Wechsler-Bellevue Subtest Scores of Three Groups Differing in Antisocial Behavior

Variable	Mean*		Variance†	*F*	*p*
Prorated IQ					
Group I	103.45	Between groups	3,296.47	9.36	.001
Group II	96.26	Within groups	352.26		
Group III	76.35				
Block Design					
Group I	9.62	Between groups	8.14	.65	Non-signif.
Group II	8.94	Within groups	12.40		
Group III	8.28				
Comprehension					
Group I	10.22	Between groups	113.06	10.77	.001
Group II	8.39	Within groups	10.49		
Group III	5.00				
Vocabulary					
Group I	9.34	Between groups	50.32	5.22	.01
Group II	8.93	Within groups	9.64		
Group III	6.56				
Similarities					
Group I	11.00	Between groups	101.72	7.47	.01
Group II	9.81	Within groups	13.61		
Group III	6.83				

* *Ns* = 24, 18, and 14 for Groups I, II, and III, respectively.
† Between groups *df* = 2; Within groups *df* = 54.

accordance with the suggestion of Hunt and French (3) would improve its discrimination along the dimension of antisocial trends.

Summary

One of the questions with which this study was concerned was whether the Porteus Maze test, qualitatively scored, could successfully distinguish between poor social adjustment marked by antisocial behavior and various other types of poor social adjustment. The second aspect of the study dealt with scatter scores among the Comprehension, Vocabulary, Similarities, and Block Design subtests of the Wechsler-Bellevue Intelligence Scale.

Fifty-six young adult Army trainees, all of them referrals to the Mental Hygiene Consultation Service, serve as *Ss*. The *Ss* were divided into three groups on the basis of wide variations in overt antisocial behavior.

Qualitative Maze Test scores differentiated significantly between three groups, the most antisocial group having the largest *Q* score. It appeared that impulsivity was the basic personality variable measured by qualitative evaluation.

Finally, our impression that Block Design performance would be the subtest least impaired by antisocial tendencies was confirmed. The results showed no significant differences between the three groups in Block Design scores, although the Comprehension, Vocabulary, and Similarities scores differed significantly. An explanation for these findings is offered in terms of the more remote examiner-subject relationship during Block Design administration as compared with the other three subtests.

Received December 20, 1955.

References

1. Docter, R. F., & Winder, C. L. Delinquent vs. nondelinquent performance on the Porteus Qualitative Maze Test. *J. consult. Psychol.*, 1954, **18**, 71–73.
2. Grajales, M. C. Porteus' Qualitative Maze Test as a measure of delinquency. Unpublished master's thesis, Fordham Univer., 1945.
3. Hunt, W. A., & French, Elizabeth G. The CVS abbreviated individual intelligence scale. *J. consult. Psychol.*, 1952, **16**, 181–186.
4. McNemar, Q. *Psychological statistics*. New York: Wiley, 1949.
5. Porteus, S. D. *Qualitative performance in the maze test*. Vineland, N. J.: Smith Printing House, 1942.
6. Porteus, S. D. Q-scores, temperament and delinquency. *J. abnorm. soc. Psychol.*, 1945, **21**, 81–103.
7. Porteus, S. D. *The Porteus Maze Test and intelligence*. Palo Alto, Calif.: Pacific Books, 1950.
8. Wechsler, D. *The measurement of adult intelligence*. (3rd Ed.) Baltimore: Williams & Wilkins, 1944.
9. Wright, Clare. The qualitative performance of delinquent boys on the Porteus Maze Test. *J. consult. Psychol.*, 1944, **8**, 24–26.

Journal of Consulting Psychology
Vol. 20, No. 5, 1956

MMPI Differences Between Parents of Stuttering and Nonstuttering Children[1]

Leonard D. Goodstein
State University of Iowa

and W. Grant Dahlstrom
University of North Carolina

While there has been much published speculation and theorizing about the etiology of stuttering (e.g., 1, 8), until recently there have been relatively few empirical studies of the onset of this fairly common clinical problem. Darley (2), Johnson (9), LaFallotte (12), and Moncur (13, 14) have all attempted to investigate the circumstances surrounding the development of stuttering by intensively interviewing the parents of stuttering children. The present study represents a cross validation and extension of these prior investigations and includes an objective measure of parental adjustment, the Minnesota Multiphasic Personality Inventory (MMPI). This paper reports the MMPI results, while the more general findings are reported elsewhere (10).

The MMPI has been used in a similar way for evaluating parental adjustment in its relation to child behavior by Williams (18) in a study of families in which there was a cerebral-palsied child, and by Grossman (7) in relation to stuttering. The present study is an extension of this method to a larger, more carefully selected set of parents.

Subjects and Procedures

A family was considered for inclusion in the stuttering or experimental group if one or both of the parents regarded the child as a stutterer and if the duration of the stuttering was reported as less than 36 months. A family was considered in the nonstuttering or control group if the child had never been regarded as a stutterer by either of the parents. Each experimental family was matched to a control family on the basis of the child's age and sex as well as the socioeconomic status of the parents. Only essentially physically normal, white children were included in the study. The experimental subjects (*Ss*) had been referred to the experimenters by speech therapists, physicians, etc., while the control *Ss* were secured by appeals for volunteers to PTA's, church groups, etc. The complete details of subject selection are available elsewhere (10).

The routine administration of the booklet form of the MMPI was part of the standard interview procedure (see 10) that was followed with each parent who had agreed to participate in the over-all study. Usually one parent took the test while the other parent was being interviewed. If the tests were not completed before the interviewer was ready to leave, arrangements were made to have the booklets and answer sheets picked up at some later time.

While it had been planned to secure an MMPI profile from each parent, this goal was not realized due to a number of uncontrollable factors such as lack of time, refusal to cooperate, etc. MMPI profiles, however, were obtained from 224 of the 300 experimental parents and from 223 of the 300 con-

[1] This research was supported by a grant from the Louis W. and Maud Hill Family Foundation of St. Paul, Minnesota. The authors wish to thank Dr. Wendell Johnson, the principal recipient of the grant, for his help and encouragement during the execution of the research and the preparation of this manuscript.

trol parents. An inspection of the other data did not suggest that those parents who had failed to complete the MMPI were different in any obvious ways from those parents for whom profiles were secured.

All MMPI answer sheets were hand scored by the interviewers for the usual four validity and nine clinical scales as well as four special scales that were thought, on an a priori basis, to represent variables important to the onset of stuttering. These four special scales were the Taylor Manifest Anxiety Scale (*A*) (15), the Achievement Scale (*Ac*) (5), the Dominance Scale (*Do*) (6), and the Status Scale (*St*) (4). All profiles were also coded into Welsh's high-point coding system (16). For ease in tabulation and statistical analysis, the scores on the 17 scales as well as the high-point codes were then coded into a predetermined coding system and punched on standard IBM punch cards.

Results

Four groups of 100 profiles each, experimental mothers, experimental fathers, control mothers, and control fathers were selected from the 447 available MMPI profiles by excluding the 47 profiles whose validity scale scores suggested a significant degree of dissimulation. The cutting scores for exclusion were: *L*-scale raw score of 7 or higher, or *F*-scale raw score of 17 or higher, or *K*-scale raw score of 28 or greater. The numbers of cases excluded on the basis of these criteria were 17, 22, and 8, respectively; there were no obvious differences between the control and experimental parents with respect to these validity criteria.

The means and standard deviations on the 17 MMPI scales for the experimental and control parents, together with the significance levels of the differences between the experimental and control means are presented in Table 1. For ease in communication the scores in Table 1 are reported in values based on the original distributions of scores rather than the coded score distributions. The reported values are, consequently, approximate rather than exact; all statistical tests, however, were based upon the more exact coded score values.

A prior comparison of the means of the experimental mothers with the experimental fathers and of the control mothers with the control fathers gave no evidence of any sex differences in the scores and the data from the two sexes were combined. The only exception to this finding was for the *Mf* scale on which both groups of fathers had significantly higher ($p < .01$) mean *T* scores than both groups of mothers; there were, however, no significant differences between the experimental and control groups and the data were therefore combined in this case as well.

There were only three significant mean differences between the experimental and the control parents on the 17 MMPI scales. The control parents had significantly higher mean scores on the *K* and Dominance (*Do*) scales ($p < .05$, $p < .01$, respectively) while the experimental parents had a significantly higher mean score on the Anxiety (*A*) scale ($p <$

Table 1

Means and *SD*s* of the MMPI Scales for the Two Groups and the Significance Levels of the Differences Between the Means ($N = 200$ in each cell)

	Experimental parents		Control parents		
Scale†	Mean	*SD*	Mean	*SD*	p‡
?	6.5	7.9	5.8	4.6	ns
L	3.3	1.9	3.6	1.8	ns
F	3.2	2.5	2.8	1.8	ns
K	15.5	5.2	16.9	4.8	.05
Hs	51.6	8.7	53.0	7.4	ns
D	53.9	10.3	52.4	9.2	ns
Hy	55.5	7.9	56.1	7.7	ns
Pd	53.2	9.6	54.8	9.0	ns
Mf§	52.5	10.1	52.7	10.1	ns
Pa	52.8	7.8	52.3	6.8	ns
Pt	53.0	8.6	51.9	7.7	ns
Sc	51.7	9.0	52.2	7.8	ns
Ma	51.9	10.1	51.2	9.1	ns
A	14.3	7.6	11.8	6.9	.01
Ac	16.9	2.9	16.8	2.6	ns
Do	17.3	3.3	18.8	3.3	.01
St	21.4	3.9	22.5	4.0	ns

* These values are approximate as they were derived from the coded scores.

† The scores on the clinical scales (*Hs*, *D*, *Hy*, *Pd*, *Mf*, *Pa*, *Pt*, *Sc*, and *Ma*) are in *T*-score units; all the others are expressed in raw score units. *K*-scale corrections included where appropriate.

‡ The *t* tests for mean differences as well as the significance levels for these differences were based upon the more precise coded score values rather than the values reported in this table.

§ Both groups of fathers were significantly higher ($p < .01$) than both groups of mothers; there were, however, no significant differences between the stutterers and nonstutterers.

Table 2

Relative Frequencies of the Initial Scale in the High-Point Code for Each Group

Group	MMPI Clinical Scales*									Total
	Hs	*D*	*Hy*	*Pd*	*Mf*	*Pa*	*Pt*	*Sc*	*Ma*	
Experimental mothers	4	13	18	12	15	12	8	6	12	100
Experimental fathers	5	17	16	8	23	4	5	4	18	100
Subtotal	9	30	34	20	38	16	13	10	30	200
Control mothers	3	18	27	22	6	7	7	5	5	100
Control fathers	1	10	17	18	27	4	5	3	15	100
Subtotal	4	28	44	40	33	11	12	8	20	200
Grand total	13	58	78	60	71	27	25	18	50	400

* Only the nine clinical scales were used in this analysis.

.01). None of the other mean differences were significant at an acceptable level of confidence.

The relative frequencies in the four groups of *Ss* of the initial scale in the high-point code, that is, the clinical scale with the highest *T*-score value in the individual profiles, are presented in Table 2. The differences between the obtained frequencies were then analyzed by means of the chi-square test for a two-by-nine table. There were no significant differences between the experimental and control fathers, between the experimental and control mothers, or between the experimental and control parents combined over sex. There were, however, significant differences between the total male group and the total female group when the experimental and control groups are combined ($p < .001$). It may be seen in Table 2 that the *Mf* and *Ma* scales occurred relatively more frequently in the initial position in the male codes while the *Pa* and *Sc* scales were the more frequent initial points in the female codes.

The relative frequencies in the four groups of *Ss* of the secondary scale in the high-point code, that is, the clinical scale with the second highest *T*-score value in the individual profiles, are presented in Table 3. The differences between the obtained frequencies were again analyzed by the means of the chi-square test for a two-by-nine table. Again there were no differences between the experimental and control fathers, the experimental and control mothers, or between the total experimental and the total control group. Again there were significant differences between the total male group and the total female group when the experimental and control groups are com-

Table 3

Relative Frequencies of the Secondary Scale in the High-Point Code for Each Group

Group	MMPI Clinical Scales*									Total
	Hs	*D*	*Hy*	*Pd*	*Mf*	*Pa*	*Pt*	*Sc*	*Ma*	
Experimental mothers	9	11	21	12	6	14	12	9	6	100
Experimental fathers	4	18	13	14	15	11	8	5	12	100
Subtotal	13	29	34	26	21	25	20	14	18	200
Control mothers	9	6	20	14	13	12	8	7	11	100
Control fathers	2	13	18	16	16	9	6	9	11	100
Subtotal	11	19	38	30	29	21	14	16	22	200
Grand total	24	48	72	56	50	46	34	30	40	400

* Only the nine clinical scales were used in this analysis.

bined ($p < .02$). It may be seen in Table 3 that the *D*, *Mf*, and *Ma* scales occurred relatively more frequently in the secondary position in the male codes while the *Hs* and *Hy* scales were the more frequent secondary scales in the female codes.

The number of parents in the two groups with one or more *T* scores above 70 on the clinical scales (the usual cutting score) was then obtained in an effort to evaluate the relative frequency of "severe disturbance" in the two groups. There were 47 persons with one or more *T* scores above 70 in the control group and 49 persons in the experimental group; by eliminating from consideration those cases where both parents in the family had scores above 70, these numbers were reduced to 41 and 42, respectively. An inspection of the high-point codes of those "severely disturbed" individuals did not reveal any differences between the experimental and control groups although the sex differences that were discussed above (see Tables 2 and 3) were again present.

In an attempt to evaluate the differences in the severity of stuttering as a function of the "abnormality" of the MMPI scores, the experimental group was then divided into two subgroups, those families in which one or both parents had one or more clinical scale *T* scores above 70 (the disturbed group) and those families in which neither parent had a clinical scale *T* score above 70 (the nondisturbed group). Independent ratings by the interviewers of the experimental children on the Iowa Scale for Rating the Severity of Stuttering (11, pp. 129–136) were available for 38 children whose parents were placed in the disturbed group and 77 children whose parents were in the nondisturbed group. The mean rating for the children of the disturbed parents was 1.97 ($SD = .97$) while the mean rating for the children of the nondisturbed parents was 3.13 ($SD = .71$); the differences between these means is highly significant ($t = 6.4$, $p < .001$) and indicates that the children of the nondisturbed parents were the *more* severe stutterers. (A rating of 1 on the Iowa Scale indicates normal, fluent speech while a rating of 3 indicates a relatively mild degree of stuttering.)

Discussion

The most striking finding of this portion of the total study was the failure to establish any major differences between the parents of the stuttering and nonstuttering children. Although certain differences between the two groups of parents were statistically significant and psychologically provocative, the over-all impression left by the results is that whatever important differences exist between these groups is not reflected in their MMPI scores. Certainly the small absolute differences between the means and the large amount of overlap between the two distributions of scores would prohibit the use of the MMPI as an instrument for differentiating these groups in clinical work.

These results are in essential agreement with those of Grossman (7) who reported that the parents of stutterers gave more atypical or pathological responses on the MMPI than the parents of nonstutterers on the F scale but did not differ on any of the other scales. Her results were obtained from older, more chronic stutterers and her parent groups involved only an *N* of 42 each. It should be noted that the differences between her two groups were small in absolute magnitude and statistically unreliable.

It is essential to note that not only are both sets of parents very similar to each other in their mean MMPI profiles but they are also very similar to the original standardization group of Minnesota normals (17). The highest clinical scale mean for either group is 56.1 for the control parents on the *Hy* scale; this represents a deviation of only 0.6 *SD*s from the expected score of a group of psychiatrically normal persons, that is, 50.0. The highest clinical scale mean for the experimental group, who presumably should be more maladjusted, is 55.5, also on the *Hy* scale. These, of course, are the largest differences from the expected normal mean score and all the other obtained means are closer to the expected value of 50.0. While some of these differences may be of statistical significance, they are not, in all probability, psychologically important. It is more reasonable to ascribe such small differences to sampling factors than to any "real" differences between the individuals studied in the present investigation and the

standardization group. These differences certainly are not evidence for regarding either the control or experimental groups of parents as psychiatrically abnormal or more maladjusted than adult persons in general.

The absence of abnormality in these mean profiles is even more strikingly brought home when these profiles are compared with the mean profiles of various neurotic and psychotic groups (17) and of college students (3, 17). Each of these latter groups have characteristic mean profiles with very marked peaks and valleys differing greatly from the flat mean profiles obtained from these parents. Such comparisons strongly support the contention that these individuals are, for the most part, essentially normal persons who differ little, if at all, from the group of normal adult Minnesotans who comprised the standardization sample. These results incidentally, especially those with the control group, offer strong support for the continued use of the MMPI as a screening instrument with presumably normal adults.

While the present results offer no evidence for concluding that these two groups of parents are very different from each other or from normal individuals generally, it should be recalled that 3 of the 17 mean differences, although the absolute differences were quite small, were significant and deserve further comment. On the *K* scale the control parents' mean was significantly higher than that of the experimental parents. This would indicate that the control parents were somewhat more defensive, less likely to admit psychological defects or problems, than the experimental parents. These control parents had not been asked, as had the experimental parents, to admit *as part of the selection procedure* that they regarded their children as deviant, and certainly might be expected to be less open and more defensive in such a clinical situation.

The finding that the control group mean on the Dominance (*Do*) scale was significantly higher than that of the experimental group, suggesting greater dominance in face-to-face personal relationships, is rather difficult to interpret and fit into the other findings and conclusions about the onset of stuttering. Little is known about the interpretation and usefulness of the *Do* scale as it is a recent addition to the MMPI scales and perhaps any more elaborate interpretation of this result should be postponed until additional information about the *Do* scale is available.

The experimental parents were significantly higher on the Anxiety (*A*) scale than were the control parents. Certainly the admission of greater tension and anxiety on the part of the stutterers' parents may be regarded as an important finding, but great care should be exercised in relating this finding to the onset of stuttering. The heightened anxiety on the part of the experimental parents may be an important etiological factor in the development of stuttering, or may be a subsequent reaction by these experimental parents to the development of stuttering by their children. The present results do not permit a resolution of these alternative interpretations.

Perhaps the best evidence that the experimental group of parents was essentially normal from a psychiatric point of view was the finding that the "deviant" group of experimental parents, those with one or more *T* scores above 70, had children with little or no stuttering as compared with the nondeviant experimental group. On the other hand, the MMPI profiles of the parents of the more severe stutterers were almost uniformly within normal limits and gave no indication of severe psychological pathology of any sort. It may be that these deviant parents volunteered for participation not so much because they were concerned with nonfluencies in their childrens' speech, but rather with their own problems; they may have regarded participation in the experiment as one way of receiving help with these more general problems, rather than an opportunity to discuss their childrens' speech. Such an interpretation would indicate a need for caution on the part of the speech therapist when he does find a stutterer's parent with a deviant MMPI profile.

The MMPI results are in general agreement with those of the over-all investigation (10) in that they do not support the notion that the parents of stutterers are very different in their behavior and attitudes from the parents of nonstutterers. Rather these results support the suggestion that those group differences that are important in the etiology of stuttering are rather limited and probably

quite specific to the speech situation and the stuttering phenomenon.

Summary

The Minnesota Multiphasic Personality Inventory (MMPI) was administered to 447 parents participating in an interview study of the onset of stuttering in an attempt to compare the relative psychological adjustment and personality characteristics of parents of stuttering children (experimental group) and parents of nonstuttering children (control group). Four groups of 100 MMPI records, experimental fathers, experimental mothers, control fathers, control mothers, were selected by excluding those records whose validity scores suggested dissimulation.

There were only three significant mean differences between the experimental and control parents on the 17 MMPI scales used in this study; the control parents had significantly higher mean scores on the *K* and *Do* scales while the experimental parents had a significantly higher mean score on the *A* scale. A comparison of the high-point codes of the experimental and control groups revealed no significant differences between these two groups. There were no significant differences in the number of parents in the two groups with one or more clinical scale *T* scores above 70, the usual cutting score indicating some degree of psychological pathology. The children of the experimental parents with *T* scores above 70 were shown to have been rated as having significantly *less* severe stuttering than the remainder of the experimental group children.

These results were interpreted as indicating that the performances of these two groups were not significantly different, nor was the performance of either group significantly different from the expected performance of psychiatrically normal persons. There was no evidence that the etiology of stuttering is rooted in gross psychological pathology of the stutterer's parents but rather the results are to be interpreted as supporting the notion that such attitudes and adjustments of the parents as may be involved in the etiology of stuttering are probably limited and specific to the speech situation.

Received January 3, 1956.

References

1. Barbara, D. A. *Stuttering. A psychodynamic approach to its understanding and treatment.* New York: Julian Press, 1954.
2. Darley, F. L. The relationship of parental attitudes and adjustments to the development of stuttering. In W. Johnson (Ed.), *Stuttering in children and adults.* Minneapolis: Univer. of Minnesota Press, 1955.
3. Goodstein, L. D. Regional differences in MMPI responses among male college students. *J. consult. Psychol.*, 1954, **18**, 437–441.
4. Gough, H. G. A new dimension of status. I: Development of a personality scale. *Amer. sociol. Rev.*, 1948, **13**, 401–409.
5. Gough, H. G. The construction of a personality scale to predict academic achievement. *J. appl. Psychol.*, 1953, **37**, 361–366.
6. Gough, H. G., McClosky, H., & Meehl, P. E. A new personality scale for dominance. *J. abnorm. soc. Psychol.*, 1951, **46**, 360–366.
7. Grossman, Donna J. A study of the parents of stuttering and nonstuttering children using the MMPI and Minnesota Scale of Parents Opinion. Unpublished master's thesis, Univer. of Wisconsin, 1951. (Abstracted in *Speech Monogr.*, 1952, **19**, 193–194.)
8. Hahn, E. F. *Stuttering. Significant theories and therapies.* Stanford, Calif.: Stanford Univer. Press, 1943.
9. Johnson, W. A study of the onset and development of stuttering. In W. Johnson (Ed.), *Stuttering in children and adults.* Minneapolis: Univer. of Minnesota Press, 1955.
10. Johnson, W. (Ed.) *The onset of stuttering.* Minneapolis: Univer. of Minnesota Press, in press.
11. Johnson, W., Darley, F. L., & Spriestersbach, D. C. *Diagnostic manual in speech correction.* New York: Harper, 1952.
12. LaFallotte, A. C. A study of the parental environment of stuttering children. Unpublished doctor's dissertation, Univer. of Denver, 1948.
13. Moncur, J. P. Environmental factors differentiating stuttering children from nonstuttering children. *Speech Monogr.*, 1951, **18**, 312–325.
14. Moncur, J. P. Parental domination in stuttering. *J. Speech Hearing Disorders*, 1952, **17**, 155–165.
15. Taylor, Janet A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.
16. Welsh, G. S. An extension of Hathaway's MMPI profile coding system. *J. consult. Psychol.*, 1948, **12**, 343–344.
17. Welsh, G. S., & Dahlstrom, W. G. *Basic MMPI readings in psychology and medicine.* Minneapolis: Univer. of Minnesota Press, in press.
18. Williams, Janet T. A study of the parents of cerebral palsied and noncerebral palsied children using the MMPI. Unpublished master's thesis, Univer. of Wisconsin, 1951.

Journal of Consulting Psychology
Vol. 20, No. 5, 1956

Response Bias and the MMPI

Eugene H. Barnes
VA Hospital, Downey, Illinois

In a recent article Berg (2) has proposed what he terms "the deviation hypothesis." This hypothesis is stated as follows: "Deviant response patterns tend to be general; hence those deviant behavior patterns which are significant for abnormality and thus regarded as symptoms, are associated with other deviant response patterns which are in noncritical areas of behavior and which are not regarded as symptoms of personality aberration" (2, p. 62). Deviant responses are defined as those which differ from modal responses or criterion group responses at a given level of significance. Berg states, "To test the hypothesis stated above, it is necessary only to identify the common or modal responses to any suitable series of stimulus patterns and predict that those subjects whose responses rather consistently go against the modal preferences will be deviant in the sense of exhibiting symptom patterns associated with abnormal states. It should be noted that content of stimulus patterns is not regarded as important" (2, pp. 62–63).

In a study pertinent to this hypothesis Barnes (1) used the Perceptual Reaction Test (PRT) as an instrument for eliciting deviant response patterns. The PRT is a series of sixty abstract designs made with ruler and compass to which the *S* is required to respond by marking one of the following options: *Like much, like slightly, dislike slightly,* or *dislike much.* By the use of deviant response patterns alone it was possible to construct scales for certain categories of psychopathology. In this study it was found that psychotic groups differed in their option preferences from normal groups in that the former generally preferred *like much* while the normal groups tended to prefer *dislike slightly.* A set to choose *dislike much* tended to be characteristic of character disorders in contrast to the performance of psychotic and normal groups. Cottle and Powell (4) demonstrated that answering all of the MMPI items "true" produced a profile with high scores on the psychotic triad while scoring all items "false" produced high scores on the neurotic triad. Berg cites these studies as evidence suggesting that, "a marked tendency to say 'yes' or 'agree,' etc., in personality inventories should be associated with psychotic states while 'no' or 'disagree' option choices should be associated with neurotic states" (2, p. 66).

There are evidently two parts to the deviation hypothesis as stated by Berg. The first and major part concerns merely the association of deviant response patterns on psychological tests with deviant behavior patterns. It asserts a general relationship between deviant test behavior and the deviant behavior of psychopathological conditions. This hypothesis draws together a variety of behavioral phenomena and makes explicit an assumption which has been implicit in much prior work. Few would argue that a "blood response" on the Rorschach was deviant behavior in a socially critical area, while few would disagree that strangling one's mother is. Nevertheless, a tenuous relationship has been assumed between "acting out" behavior and the socially nonsignificant response of "blood." The second and minor portion of the hypothesis concerns the directionality of such deviant response patterns as have been observed. This portion of the hypothesis asserts, in effect, that different types of response sets or biases are associated with different types of deviant behavior.

Individual record forms of the MMPI are a convenient source of material for testing

certain deductions from the deviation hypothesis since the scoring system provides that only those items which are answered in the infrequent direction are recorded (6, p. 8). These items are, by Berg's definition, deviant responses. Thus, if content of stimulus patterns is not important, it would be inferred that the number of deviant responses, i.e., *X*s, on the record form of the MMPI should be as effective a measure of psychopathology in general as the profile itself. A scale of the MMPI which specifically purports to measure a psychotic reaction is the *Sc* scale. This scale, then, should correlate significantly with the total number of deviant responses. A preliminary check with grouped data revealed that the number of *X*s correlated .92 with the *Sc* scale for a sample of 24 male veterans in a neuropsychiatric hospital. The correlation to be expected from the number of common elements is .38 as calculated from a formula given by McNemar (8, p. 118). This limited confirmation suggested a more systematic exploration.

A large number of studies have indicated that the MMPI and the groups of scales known as the neurotic triad (*Hs, D, Hy*) and psychotic triad (*Pa, Pt, Sc*) have useful clinical validity (3). By inference from the deviation hypothesis, the total number of deviant responses should generally correlate with scales indicative of "symptom patterns associated with abnormal states" (2, p. 63). From the second part of the hypothesis it follows that deviant responses on the MMPI resulting from responding "true" when the modal response is "false" should correlate most highly with the scales of the psychotic triad while responding "false" when the modal response is "true" should correlate more highly with the scales of the neurotic triad. Since these tendencies to prefer certain option choices are conceived as being different response sets, the tendency to answer atypically true should not be correlated with the tendency to answer atypically false.

Procedure

Subjects. The *S*s were 40 males who, for one reason or another, had been given the MMPI by the author. These *S*s were a very heterogeneous group including psychotic outpatients, neurotics, vocational counseling cases, and court referrals. Variances on the MMPI clinical scales are, therefore, doubtless larger than from a random sample of the general population. It should also be noted that they were not representative of any homogeneous clinical grouping. The group was probably representative of those who come to the attention of a psychologist in private practice.

Method. Pearson *r*s were machine calculated for correlations of each of the clinical scales and three validating scales with, first, total number of deviant responses, second, number of deviant responses resulting from answering items false, and third, number of deviant responses resulting from answering items true.

Results

Results are summarized in Table 1. It will be noted that the differences in relative sizes of correlation coefficients on the neurotic triad and psychotic triad are all in the predicted direction, a result significant, without respect to magnitude, at less than a 2 per cent level of confidence as calculated by the sign test. The correlation between the deviant responses obtained by answering false and those obtained by answering true is less than the standard error and we may accept the hy-

Table 1

Correlations of MMPI Scales with Deviant Responses

Scale	Total number of deviant responses	Deviant responses answered true	Deviant responses answered false
L	−.09	−.24	.37
F	.74	.66	.43
K	−.60	−.69	.06
Hs	.66	.52	.56
D	.56	.34	.73
Hy	.12	−.09	.59
Pd	.71	.68	.28
Mf	.32	.28	.21
Pa	.60	.52	.37
Pt	.87	.84	.31
Sc	.93	.90	.36
Ma	.38	.50	.21
Dev. responses answered false		.11	

pothesis of no correlation. Certainly the total number of deviant responses is significantly associated with the clinical scales.

Discussion

This study was designed to test certain deductions derived from the deviation hypothesis and collateral research, not as a definitive test of the total hypothesis. Within this restricted frame of reference the deductions have been confirmed and this confirmation indirectly lends support to the hypothesis from which they were derived.

The deviation hypothesis does not take into consideration the matter of item content. If a preference for a response option exists in a normal population, deviating from this preference should add some increment to an indicator of psychopathology. In the present study we find evidence of such a phenomenon. With no attention given to particular items or their content, a mere summing of the deviant responses bears a striking relationship to certain of the MMPI clinical scales, as shown in Table 1.

One interesting implication of this portion of the hypothesis is that it would provide a theoretical rationale for the behavior of the *F* scale of the MMPI. This scale was constructed from items which were very rarely answered in the scored direction by normal *Ss* (5). On inspection these items seem to have no psychological coherence or communality, yet it has been observed that the *F*-scale score is elevated in personality aberrations (7, 9). The *F* scale has even been dubbed somewhat informally and inelegantly "the screwball scale."[1] The deviation hypothesis would predict just such a result from a scale so constructed. It would, in fact, predict that even a random selection of items scored in this manner would give a similar result.

With respect to the direction of deviation, it is clear that deviating by atypically answering true is associated with MMPI scales indicative of a psychotic type of pathology while atypically answering false is associated with scales indicative of a neurotic reaction. It may be argued that this result is obtained simply because the scales of the neurotic triad are more heavily weighted with items scored when they are answered false and those of the psychotic scales more heavily weighted with items scored when answered true. Although it is a fact that a degree of such differential weighting exists, such an "explanation" is circular in that it fails to explain the origin of such a weighting. Why did it happen, in the empirical derivation of the scales, that more items were found to discriminate between certain clinical groups and normals when answered in just these directions? In view of the results obtained in this study and the other studies cited, an answer in terms of chance distribution of item content seems untenable. It seems increasingly probable that response sets play a much more important role in personality testing than has been suspected.

[1] If memory serves correctly the author first heard this label used by Dr. E. S. Bordin.

Summary

A study was undertaken to test some deductions concerning the MMPI derived from Berg's deviation hypothesis (2). According to this hypothesis *Ss* whose responses consistently go against modal preferences should display symptom patterns of abnormality. Further, this hypothesis would predict that atypically responding true should be associated with psychotic states while atypically answering false should be associated with neurotic conditions. These two tendencies should be relatively unrelated to one another.

Since the record form of the individual MMPI provides that only items answered in the infrequent direction are entered, it is possible to count the number of deviant responses, to separate these into those obtained by answering atypically false and atypically true, and to correlate these scores with the scales of the MMPI. These operations were performed and it was found, as predicted, that the total number of deviant responses correlated highly with the clinical scales, that the tendency to answer atypically true was more highly correlated with the scales of the psychotic triad while the tendency to answer atypically false was more highly correlated with the neurotic triad, and

that the two tendencies were not significantly correlated.

It is concluded that the deviation hypothesis has received some indirect confirmation both with respect to its general statement and with respect to direction of deviation. Also, it is concluded that response sets of this type play a larger part in personality testing than has been suspected.

Received December 1, 1955.

References

1. Barnes, E. H. The relationship of biased test responses to psychopathology. *J. abnorm. soc. Psychol.*, 1955, **51**, 286–290.
2. Berg, I. A. Response bias and personality: The deviation hypothesis. *J. Psychol.*, 1955, **40**, 61–72.
3. Cottle, W. C. The MMPI: A review. *Kan. Stud. Educ.*, 1953, **3**, No. 2.
4. Cottle, W. C., & Powell, J. O. The effect of random answers to the MMPI. *Educ. psychol. Measmt*, 1951, **11**, 224–227.
5. Hathaway, S. R., & McKinley, J. C. *Manual for the Minnesota Multiphasic Personality Inventory*. (Rev. Ed.) New York: Psychological Corp., 1943.
6. Hathaway, S. R., & McKinley, J. C. *Minnesota Multiphasic Personality Inventory:* Manual. (Rev. Ed.) New York: Psychological Corp., 1951.
7. Kazan, A. T., & Sheinberg, I. M. Clinical note on the significance of the validity score (*F*) in the MMPI. *Amer. J. Psychiat.*, 1945, **102**, 181–183.
8. McNemar, Q. *Psychological statistics*. New York: Wiley, 1949.
9. Schmidt, H. O. Test profiles as a diagnostic aid: The MMPI. *J. appl. Psychol.*, 1945, **29**, 115–131.

Journal of Consulting Psychology
Vol. 20, No. 5, 1956

Delaying Capacity, Fantasy, and Planning Ability: A Factorial Study of Some Basic Ego Functions

Jerome L. Singer, Harold Wilensky, and Vivian G. McCraven
Franklin D. Roosevelt VA Hospital, Montrose, N. Y.[1]

An increasing ability to delay gratification and to inhibit a direct motor approach to a desired object appears to be a necessary feature in the development of an identifiable self in human personality. Intimately associated with this delaying capacity is the ability to resort to thought or ideation as "experimental action" (8) in a behavioral medium freed of temporal and spatial boundaries. As early as 1900 Freud (7) had suggested a theoretical linkage of delay and thought development and this relationship became a pivotal feature of his concept of the reality principle and, later the ego. Similar theoretical and empirical associations of thought and motor inhibition or delay have also been described by psychologists and psychiatrists in diverse fields (13, 19, 21, 30). With the increased attention paid by psychoanalytic theorists to ego psychology and to functions such as planning, synthesizing, and controlling motility (9, 11, 19), the possibility of integrating the psychoanalytic hypotheses with the concepts of psychologists involving the theoretical link between fantasy, language, and motor behavior emerges more clearly (22). The present study represents a further step in a program of research aimed at investigating some relationships of thought and delay in ego structure.

The specific approach to this problem has been through the medium of an intensive study of the Human Movement response (*M*) on the Rorschach inkblots and its behavioral correlates. Rorschach (20) pointed out that those persons who saw numerous human figures in motion on the blots tended to be inhibited in their own motor activity and given to considerable imagination or fantasy, "movements in the mind's eye." In effect, therefore, Rorschach's observations, and those of subsequent Rorschach workers (16, 18, 21) suggest that the *M* response may serve as a crude measure of delaying capacity and of imaginative tendencies which are basic ego-functions in Freud's formulation. The experimental work on the vicarious functioning of sensory and motor patterns carried out by Werner and Wapner and their colleagues is consistent with this approach and has carried it a step toward more precise formulation in perceptual theory (30, 31, 32). A number of studies in the present series have also supported Rorschach's observations by demonstrating that individuals whose Rorschachs contain relatively more *M* responses are capable of greater motor inhibition in slow-writing tasks, move about less in a waiting-room situation, and tend to tell more imaginative stories to TAT pictures (14, 23, 24). In the study to be presented here this work will be extended to include a large number of tests and behavioral situations which presumably tap aspects of impulse control or delaying capacity, planfulness, and fantasy. The general problem raised is whether consistent individual patterns emerge along a dimension that links delaying capacity, fantasy, planning ability, and various measures of motility control.

In approaching a concrete formulation of the general hypothesis a series of tests and behavioral indices were selected which, while

[1] The authors are indebted to Drs. John F. Winne and Jacob Cohen who gave valuable assistance in the statistical analysis of the data and to the psychology staff, trainees, and ward personnel who participated in some of the evaluative procedures.

superficially dissimilar, were expected to define a common fantasy-motility factor. The tests included a general intelligence measure, employed as a control since it was felt necessary to determine the extent to which the dimension in question was related to intelligence. The Rorschach test was included, with Rorschach's *M* response in particular expected to yield a high loading on a fantasy-motility factor. In order to evaluate the generality of the Human Movement phenomenon and to control for total number of Rorschach responses, Barron's Movement Threshold Inkblots (1) were included. Another estimate of fantasy tendencies was obtained from stories told by the *Ss* to four Thematic Apperception Test pictures scored for "transcendence," i.e., introduction of characters, emotions, and temporal references which went beyond the mere description of the picture. Still another manifestation of fantasy tendencies characteristic of ego functioning might be an inclination toward setting high goals. A measure of level of aspiration was therefore included as well as a measure of sheer speed of performance in a simple motor task. The latter measure, however, would not be expected to show a significant loading on the factor in question. Another aspect of the adaptive nature of a developed capacity for fantasy was considered to be the ability to admit to awareness a painful experience. Some qualitative results in a previous study (33) had suggested interesting individual differences in admission of distress by *Ss* following a failure experience which had measurably interfered with intellectual performance. A crude measure of admission or denial of disturbance following failure was therefore included in the test battery with the hypothesis that it would carry an appreciable loading on a fantasy factor. A somewhat similar rationale led to the inclusion of a technique which was felt to measure compliance with the examiner's wishes and persistence in a routine task to please another person despite psychological satiation. Finally, some attempt was made to obtain a measure of planning ability or foresight since this seems so clearly a characteristic of a well-developed fantasy-life. One approach involved self-ratings on a simple scale devised by Small (26). The other called for use of the Porteus Mazes, long described as measures of foresight by their originator (17), a claim afforded at least some support in a recent study by Small (26). The Porteus Mazes would appear particularly relevant since success in their execution involves the capacity both to delay motor reactivity, curbing impulsive solutions, as well as the more ideational component of visually moving through the mazes in advance of the pencil.

Turning now to tasks and situations involving delaying capacity and control of motility, one possibility that suggested itself was a time-estimation procedure. Since the large majority of *Ss* tend to respond too soon when instructed to indicate when a stated time interval has elapsed, sheer accuracy gave an adequate measure of delaying ability and control of impusive reaction. A better measure might have been to consider excessive delaying, but this occurred too rarely in prior trials of the procedure. A measure of motor inhibition, a slow-writing task originally devised by Downey, was included. A number of previous studies had suggested a consistently positive relationship between the ability to delay writing tempo and Rorschach's *M* response (14, 23, 24). To obtain some evaluation of longitudinal gross motor tendencies, ratings on certain variables were obtained from ward personnel who knew the *Ss* for at least three months. Four variables based on the ward behavior items drawn from the Multidimensional Rating Scales (12) were employed. One involved aggressive behavior, frequently acting out a hostile impulse either through physical assaults or verbal imprecations. Another, cooperativeness with ward routine, presumably an example of controlled motility, would be expected to show a positive loading on a fantasy-motility factor. A measure of external interests, as manifested by behavior demonstrating some measure of orientation toward the hospital activities program and other people, was included. Finally, a series of items dealing with diffuse energy expenditure, gross motor activity not evidently hostile in intent, was employed, which might be expected to yield a negative or zero loading on a fantasy-motility factor.

In summary, then, the specific form of the

investigation involved the prediction that a battery of procedures which on the surface measured a wide variety of performances ranging from fantasy capacity as seen in storytelling, perception of humans in activity on inkblots, planning in a maze test, slow-writing, and controlled motility, would all be revealed as aspects of an underlying, more general dimension in which the capacity for imaginal living and control of impulsive motility were linked. Since clinical experience and some previous research had suggested the value of Rorschach's Human Movement response as a measure of both the fantasy and motility aspects of the assumed underlying factor, it was anticipated that Rorschach's *M* response and the threshold for Human Movement on Barron's inkblots would show especially high loadings on this fantasy-motility factor.

Method

Subjects. There were 100 *Ss*, all male veteran patients bearing the diagnosis of schizophrenia. They were chosen chiefly on the basis of their having recently received a Rorschach examination and an intelligence test. Their average age was approximately 31, mean date of first hospitalization 1949, mean IQ 108, and their years of education averaged 10.5. Social class background for this group was predominantly lower-middle and upper-lower and their ethnic backgrounds approximated the general hospital picture. Although all the patients in this study were unquestionably psychotic with hallucinations or delusional trends, only those *Ss* who proved to be in sufficient contact to follow instructions for all of the tests were employed in the sample. It should be noted that the patients in this study represented a different sample of *Ss* from those used in two prior investigations in this series (23, 24).

Procedure. A battery of special procedures was organized with an instruction booklet, and a scoring sheet.[2] With the exception of the Rorschach and Wechsler-Bellevue, which had usually been given some weeks or occasionally months earlier, all the data were obtained at one 90-minute sitting. The behavior ratings were obtained shortly after testing by interviews with ward personnel. Much of the testing was carried out by two of the authors (J. L. S. and H. W.) with occasional administrations of the battery by clinical psychologist trainees who had received instruction in the testing procedures.

Wechsler-Bellevue Intelligence Scale (*Variable 1*). For various practical reasons it was not possible to obtain full-scale IQs on all *Ss*, but a prorated IQ estimate was obtained. These estimates were based on verbal tests usually, and only occasionally were performance test IQs available.

Rorschach. The Rorschachs were standardly administered examinations scored in accordance with Klopfer and Kelley (10). All protocols were rechecked by conference to eliminate the few questionable scorings of the original examiners. The remaining test battery was administered by the investigators without prior knowledge of Rorschach data for the *Ss* in order to eliminate unconscious influence by the examiner. All determinants were scored but only those with sufficient frequency of occurrence and range in the sample were employed in the analysis. The Rorschach scores with Variable Numbers were: Total Responses (*R*, var. 2), Human Movement (*M*, var. 3), Animal Movement (*FM*, var. 4), Form-level (Beck's $F+\%$, var. 5), Black or Gray shading (Klopfer's *C'*, Beck's *Y*, var. 6), Form-Color (*FC*, var. 7), and Color-Form (*CF*, var. 8).

Barron's Movement Threshold Inkblots (*Variable 9*). This test consists of thirty black blots presented one at a time for thirty seconds with usual Rorschach instructions. Only one response per card is obtained and the score is based on the card in which the first clear *M* response is elicited. For purposes of simplifying the factor analysis, the scores on this variable were arbitrarily reflected prior to insertion in the matrix; a positive correlation in the matrix thus means that a low Movement Treshold is associated with a high score on the other variable.

Thematic Apperception Test Transcendence Index (*Variable 10*). Four cards of the TAT were administered in accordance with instructions developed in an earlier study (24). The cards were 1, 6BM, 7BM, and 12BM. The Transcendence Index, adapted from work by Weisskopf (29) called for scoring of items that involved addition of material not directly on

[2] Copies of the correlation matrix are available on request from the authors. Tables giving the intercorrelation matrix, the transformation matrix, instructions for the various test techniques, and sample data sheet have been deposited with the American Documentation Institute. Order Document No. 4968 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

the card, e.g., new characters, names, past or future events, etc.

Planning Self-Rating (*Variable 11*). This simple six-point rating scale was used by Small (26) in a study of planning tests. It calls for *S* to rate himself along a scale of planfulness in which each point is defined. The points range from: "I never plan things in advance," through "I always plan things out carefully in advance."

Porteus Mazes (*Variable 12*). This test was given in accordance with instructions in Porteus' Manual (17). Only the quantitative score (TQ) was employed here although further analysis by qualitative means is being planned.

Wechsler's Number-Square Test (*Level of Aspiration*). This procedure, modified from a technique devised by David Wechsler, called for *S* to draw squares around numbers printed serially in rows of ten across a page. Instructions called for speed and 30 seconds were allotted for a trial of 40 numbers. Prior to the first trial *Ss* were asked to estimate their expected performance given only the information that "people do between 18 and 30 in half-a-minute." Following performance a second estimate was obtained. This procedure was carried out twice so that there were in all three expectation scores (one prior to performance) and two actual performance scores. Measures included Initial Expectation score (E_1, Var. 13) and Level of Aspiration $E_2 + E_3 - P_1 + P_2$ (L.A., Var. 15). The score obtained on initial performance (P_1, Var. 14) was also employed as a measure of sheer speed of reaction.

Time Estimation (*Variable 16*). This task called for *S* to indicate by saying "Stop" when he thought that a given time period had elapsed. Three times were used, 15, 30, and 60 seconds. Deviations from actual time were added to give a total deviation score. The measure used was *Accuracy*, based on absolute deviation.

Block Authority-Reaction Test (*Variable 17*). This procedure modified from a technique devised by the Blocks (2), called for *Ss* to put ten pegs in holes and remove them at their own pace, repeating this procedure until their psychological satiation point was reached. The *Ss* were informed that they could stop at will, but when they did so *E* asked, "Wouldn't you like to do some more?" and then, "Are you sure you'd like to stop now?" If *Ss* continued the task in this ambiguous situation, they were scored for positive response to authority.

Motor Inhibition Test (*Variable 18*). This technique, derived originally from the Downey Will-Temperament Scale, had been used in a number of studies in this series (14, 23, 24). It called for *S* to write the phrase, "New Jersey Chamber of Commerce," as slowly as possible without stopping motion of the pencil. The score was the time in seconds taken to complete the phrase. The difference between an initial trial at normal tempo and the experimental trial was the actual measure.

Digit Frustration, Admission of Disturbance (*Variable 19*). This task, employed originally in a study on frustration by Wilensky (33), involved straightforward digit retention except that *S* was taken beyond his limit of success, and his errors on the digit series beyond his memory span pointed out. Following a standard number of failures, the digit spans were presented in decreasing length until a long period of sustained success occurred. Evidence of disturbance was based on the difference between initial level and postfrustration level of digit retention. The variable employed here was the admission or denial of disturbance following failure. This was a dichotomous classification; those patients whose responses were unintelligible or who actually refused to respond were not considered, and consequently the *N* for this variable was lower than for the others.

Behavior Ratings. To obtain some evaluation of long-term motility patterns, items from the Multidimensional Rating Scales were grouped under four headings: those pertaining to *Aggression* (Var. 20), verbal or physical assaultive and destructive behavior on the ward, *Cooperativeness* (Var. 21) with ward routine, *Interest Level* (Var. 22), as seen in voluntary participation in ward activities, and *Diffuse Energy Level* (Var. 23), excessive motor activity not specifically hostile in intent. Ward personnel, aides, nurses, physicians, and ward clinical psychologists, all of whom had daily contact with the patients in question as well as with the hospital's daily listing of unusual behavior on each ward, were interviewed and each patient rated on a three-point scale for each item in the four categories. The rater was not cognizant of test results at the time of his interviews and ratings.

Results and Statistical Analysis

Following inspection of the various distributions, the scores on each variable were dichotomized at the median. In almost all cases good splits were obtained, the exceptions being the Authority-Reaction, where only 30% of *Ss* met the criterion of compliance and the Energy-Level variable, where the cut was hampered by an extremely restricted range of scores. Following dichotomization, phi coefficients were obtained and transformed to estimates of Pearson's *r* in accordance with the procedure suggested by Edwards (5). A comparative sampling of tetrachoric coefficients revealed no sizable differences in the coefficients obtained by the two procedures—most deviations being of the order of 0.01 to 0.03. Recognizing the limitations of *r*s obtained in this fashion, the authors are following Wittenborn's reasoning (34) in attempting a factor-analytic study of Rorschach data.

A cluster analysis had led to the elimination of a number of variables which showed

Table 1

Centroid Loadings, Communalities, and Common Factor Loadings

Variables	I	II	III	IV	h^2	A	B	C	D
1. IQ (W-B)	27	26	−06	−17	17	01	25	19	−12
Rorschach Variable									
2. *R*	72	27	21	18	66	07	35	53	44
3. *M*	82	−67	25	−02	1.18	81	10	−27	65
4. *FM*	52	24	32	32	53	−07	27	52	52
5. *F*+%	32	−11	11	−17	16	22	21	−08	14
6. *C′*	25	55	−19	10	41	−19	14	57	−14
7. *FC*	46	29	−16	33	43	07	−03	56	15
8. *CF*	22	18	−16	28	18	02	−11	37	06
Experimental Battery									
9. Mt	60	−47	14	33	71	56	−14	01	62
10. TAT	58	27	30	06	50	−03	44	42	38
11. PSR	38	28	08	−47	44	02	55	07	−13
12. Port	45	−12	−28	−24	36	48	03	−07	−13
13. E_1	38	16	38	−52	58	−01	71	−06	09
14. P_1	31	42	21	−41	48	−18	64	18	−07
15. LA	11	−30	13	−11	14	23	04	−27	15
16. TIM	35	12	−18	12	18	17	−03	28	02
17. AR	24	−16	39	30	33	04	07	11	57
18. MIT	74	−25	−23	−10	67	69	00	00	12
19. AD	45	−26	36	20	44	24	12	05	60
20. Agg	19	36	−16	17	22	−11	02	44	−06
21. Coop	19	−54	−26	−34	52	64	−16	−53	−15
22. Ex I	28	27	−50	−49	64	23	16	04	−59
23. En L	15	33	−29	07	22	−04	−03	35	−21

negligible correlations in the matrix, e.g., Age, Date of First Hospitalization, Education (which correlated chiefly with IQ). The resulting matrix of 23 variables was then subjected to a factor analysis, using Thurstone's centroid method (27). Communalities were estimated by means of minor centroids. The relatively small residuals remaining in the third residual matrix and the unknown reliability of the method of estimating *r* warranted extraction of only four factors. The axes were rotated to oblique simple structure following Thurstone's method of rotation by two-dimensional sections. The centroid factors, the calculated communalities and oblique factor loadings are presented in Table 1. Since the oblique factors show slight intercorrelations, a second-order factor analysis was carried out. The matrix of intercorrelations of the factors, the second-order centroid factor loadings, and the rotated factor loadings are presented in Table 2.

Consideration of Factors A and D suggests at least partial support for the hypothesis of the study since Rorschach *M* and Movement-Threshold show sizable loadings on the two factors in which measures of controlled motility and fantasy are also appreciably loaded. Factor A derives by far its highest loadings from *M*, Movement-Threshold, Porteus Mazes, Motor Inhibition, and Cooperation in Ward Behavior. It thus appears to reflect controlled motility primarily and it seems possible that it represents the kind of motor restraint that would lead to compliance and inhibition of impulses. Factor D, on the other hand, with its sizable loadings on *M*, Movement Threshold, Admission of

Table 2

Intercorrelations Among Factors, Second-Order Centroid Loadings, and Oblique Rotated Loadings

Factor	A	B	C	D	I	II	h^2	AA	BB
A		40	44	−15	68	−16	49	15	45
B			12	−17	51	−39	41	−14	58
C				−22	58	29	42	50	00
D					−36	−19	17	−32	−01

Distress after Failure, Animal Movement, External Interests (negative loading), Continuation of Task in Response to Authority, Rorschach Responses, and Transcendence on TAT stories, appears to represent an ideational or introspective dimension of personality, particularly since all of the tasks involve little or no gross motor activity. Factor B might well be termed an ambition or need-achievement factor since it combines high loadings on productivity measures (Performance Speed, Rorschach Responses, and TAT Transcendence) with self-estimates of achievement potential (Initial Expectation on Number Square Test) and planfulness (Planning Self-Rating). Factor C with its high loading on Rorschach Color and Shading, Animal Movement, Rorschach Responses, Cooperation (Negative), Aggressive Ward Behavior, TAT Transcendence, and Energy Level in Ward Behavior, seems to represent a combination of productivity or drive with emotionality and uninhibited activity. The high loading of Rorschach *C'* and *FM* would appear to support certain views of Piotrowski (16), who has suggested that numerous *C'* and *FM* responses in Rorschach records are indicative of acting-out of impulses and animation, respectively. Factor C may well represent some aspect of a pressure for outward direction of living, perhaps a flight into activity and environmental manipulation which is manifested both motorically (high Aggression and Energy Level, low Cooperation) and ideationally (high TAT Transcendence and Rorschach Responses). It closely resembles a factor termed "Emotional Surgency" by Foster (6) [3] in a somewhat similar study of Rorschach data.

In the second-order analysis, Factor AA has a high positive loading for *Emotional Surgency* and a moderate negative loading on *Introspectiveness*. This result suggests at least some validity to Rorschach's original concept of the experience type. Factor BB is clearly defined by high loadings in *Controlled Motility* and *Ambitiousness* or *Need-Achievement*.

[3] A recent paper by Foster (6) which came to the attention of the authors subsequent to the completion of this study also describes a factor termed "emotional surgency" which has high loadings for Rorschach shading and color responses.

Discussion

The results of the present investigation, while grossly supporting the formulation advanced at the outset, make it clear that, at least for this sample, the initial hypothesis of a fantasy-motility dimension was overly simplified. The fact that Rorschach's Human Movement response showed significant correlations with almost all of the variables presumably measuring fantasy, planning or delaying ability, and motor inhibition is striking confirmation of the original formulation of Rorschach (20) concerning Human Movement and of subsequent interpretations of its theoretical meaning and behavioral implications (14, 18, 22, 30). Although Rorschach *M* and Barron's Movement-Threshold show high loadings on both Factors A and D, the emergence of two separate and relatively uncorrelated factors, one primarily involving motor control and the other ideational activity, suggests a more complex situation than was hypothesized. It seems possible that at least two types of persons are likely to present Human Movement responses to inkblots. One type consists of persons with considerable ability to restrain impulsive response and to plan ahead and cooperate for their own good in the numerous hospital situations involving waiting or complying with routine. The other group, while sensitive to the wishes of authority and perhaps willing to comply, is too withdrawn into introspection and self-evaluation or autistic reverie to make effective use of their potentialities for effective motor control and planfulness. This speculative distinction suggests more extensive qualitative exploration of the nature of Movement responses, perhaps to determine whether differences in content or elaboration of the percept may account for the association of *M* responses with the two separate patterns of response. Inasmuch as the two Human Movement measures have high loadings on both the Motor and Ideational factors in this study, the theory linking movement perception with controlled motility and fantasy receives some support. It may well be that the split factors which have emerged

may be a consequence of the pathology of the patients and that study of a normal sample might clarify this issue considerably. Should fantasy-motility linkages emerge in other studies of this sort, the possibility of an operational formulation of the concept of ego functions may open important research avenues. At a theoretical level, it may become feasible to explore the nature of individual differences in thinking and motility as a function of family constellation or cultural background (25). From a practical standpoint the development of a battery of tests of the sort described here may afford an opportunity to improve the nature of individual psychological testing by the use of brief tests tapping separate aspects of a common function. Such a development has indeed been advocated by investigators like Zubin (35), who have been critical of the interpretative complexities inherent in the use of global projective devices like the Rorschach.

The two remaining factors which emerged in this study, while not anticipated, provide interesting possibilities for fruitful study as well. *Emotional Surgency,* with its loadings on ideational productivity, aggressive or diffuse motor behavior, and Rorschach Color and Shading, calls attention to the possibilities for exploring still another dimension of perception and action. It is possible that some element of lack of perceptual control (34) goes along with a concomitant inability to restrain motor behavior. This hypothesis would not sufficiently explain the high loading of Rorschach Animal Movement responses and of Form-dominated Color responses in this sample, however, and it remains to be seen whether further studies duplicate these results. Factor B, tentatively termed *Ambitiousness* or *Need-Achievement,* suggests the picture of individuals who feel compelled to assert that they are planful and can do well in a task they have not attempted before and press hard for such accomplishment. It may be important to note that IQ, which does not load appreciably on any of the factors, has its highest value in Factor B. Thus, at least for this sample of schizophrenics, it appears that a personality dimension involving achievement drive and a need to appear planful may be delineated and proves to be relatively independent of fantasy tendencies.

A final question that may be raised in connection with these findings of the second-order analysis is whether the four factors bear any diagnostic psychopathological significance. One might, for example, speculate on a hierarchical ordering of Factors BB and AA relative to an ideal criterion of mental health (or well-integrated ego functioning). Thus, Factor BB with its pattern of planfulness, motor control, and ward cooperativeness as well as ambitiousness, might represent a more effective level of functioning than Factor AA, where there is a suggestion of excessive withdrawal (Low External Interests) and Emotional Surgency with its indication of aggressive behavior and uncooperativeness as well as diffuse motor activity. Since duration of illness and florid symptomatology did not appear to differentiate between various subgroups in this sample, such an ordering of psychopathological tendencies seems less tenable although it remains reasonable to expect that an optimal association of fantasy and delaying capacity would lead to more effective living. For the present these results, like those of Barron (1), suggest that differing styles or patterns of fantasy and motility behavior are indeed discriminable through the use of relatively simple psychological testing procedures. It seems clear that many of the problems raised by consideration of these results are relevant for general personality theory and indicate the value of research that is not geared toward differentiating one dubious psychiatric nosological group from another, but rather toward clarifying the interrelations of certain key human functions as the basis for what may ultimately become a more meaningful and objective system of personality classification. That the delineation of functional personality dimensions is not purely speculative is suggested by the fact that Factors A, D, and C here bear some resemblance to factors reported by Foster (6) as "Delay" and "Emotional Surgency," by Cattell (4) as "Nervous, alert reactivity" and "Cultured introspective self-control," as well as by Thurstone (28) as "Reflectiveness" and "Impulsivity." The work of Block and Martin (3) in delineating "Under-" and "Over-control-

ling" children seems also to bear directly on these factors.

Summary and Conclusions

A number of concepts drawn from the psychoanalytic theory of thought development, Werner's sensory-tonic theory of perception, and Rorschach's work on the Human Movement response (*M*), have led to the postulation of a dimension of personality which involves a linkage of fantasy tendencies with control of motility and impulsive behavior. In order to test this hypothesis a battery of tests which presumably measured fantasy tendencies, level of aspiration, planning ability, delaying capacity, motor inhibition and motility, and responsiveness in ward behavior was developed. The battery was administered to 100 male schizophrenic patients. A factor analysis yielded four centroid factors which were subjected to an oblique rotation to simple structure. The four rotated factors were tentatively designated as:

1. Motor Inhibition and Planfulness (high loadings on Rorschach *M*, Movement-Threshold, Porteus Maze Test Quotient, Motor Inhibition Time, and Cooperativeness in ward behavior); *2. Ambitiousness or Need-Achievement* (high loadings on Initial Expectation in a level-of-aspiration situation, Initial Performance in the same task, Planning Self-Rating, Transcendence in TAT stories, and Rorschach Response Total); *3. Emotional Surgency* (high loadings on Rorschach Responses, Animal Movement, Shading and Color, TAT Transcendence, Aggressiveness on the ward, Uncooperativeness, and Diffuse Motor Activity); *4. Introspectiveness or Introversiveness* (high loadings on Rorschach *M*, Animal Movement Responses, Movement Threshold, TAT Transcendence, Admission of Distress following Failure, Continuation of Task in Compliance with Authority, and Lack of External Interests in Ward Behavior).

The high loadings of Movement-threshold and Rorschach's *M* on both the Motor Inhibition and the Introspective Factors suggest partial support for the original hypothesis but the emergence of two separate factors for motility and fantasy activity indicates that the original formulation was overly simplified. The *Emotional Surgency* factor affords some support to Piotrowski's analysis of the behavioral implications of Rorschach Shading and Animal Movement responses. Two second-order factors, one linking Emotional Surgency and lack of Introspection and the other Motor Inhibition and Ambitiousness also emerged.

Received December 8, 1955.

References

1. Barron, F. Threshold for the perception of human movement in inkblots. *J. consult. Psychol.*, 1955, **19**, 33–38.
2. Block, J., & Block, Jeanne. An interpersonal experiment on reactions to authority. *Hum. Relat.*, 1952, **5**, 91–98.
3. Block, Jeanne, & Martin, B. Predicting the behavior of children under frustration. *J. abnorm. soc. Psychol.*, 1955, **51**, 281–285.
4. Cattell, R. B. The principal replicated factors discovered in objective personality tests. *J. abnorm. soc. Psychol.*, 1955, **50**, 291–314.
5. Edwards, A. *Statistical analysis for students in psychology and education.* New York: Rinehart, 1946.
6. Foster, A. The factorial structure of the Rorschach test. *Texas Rep. Biol. Med.*, 1955, **13**, 34–61.
7. Freud, S. The interpretation of dreams. In *The standard edition of the complete psychological works,* Vol. V. London: Hogarth, 1953. Pp. 339–628.
8. Freud, S. Formulations regarding the two principles in mental functioning. In *Collected Papers,* Vol. IV. London: Hogarth, 1946. Pp. 13–21.
9. Hartmann, H. Comments on the psychoanalytic theory of the ego. In *The psychoanalytic study of the child,* Vol. V. New York: International Universities Press, 1950. Pp. 74–96.
10. Klopfer, B., & Kelley, D. M. *The Rorschach technique.* Yonkers, N. Y.: World Book Co., 1942.
11. Kris, E. The development of ego-psychology. *Samiksa,* 1951, **5**, 153–168.
12. Lorr, M. Multidimensional scale for rating psychiatric patients. Hospital Form. *VA Tech. Bull.*, Nov. 16, 1953, No. 10–507.
13. Luria, A. P. *The nature of human conflicts.* New York: Liveright, 1932.
14. Meltzoff, J., Singer, J. L., & Korchin, S. J. Motor inhibition and Rorschach movement responses: a test of the sensory-tonic theory. *J. Pers.*, 1953, **21**, 400–410.
15. Phillips, L., & Smith, J. G. *Rorschach interpretation: Advanced technique.* New York: Grune & Stratton, 1953.

16. Piotrowski, Z. A. A Rorschach compendium, revised and enlarged. *Psychiat. Quart.*, 1950, **24**, 543–596.
17. Porteus, S. D. *The Porteus Maze test and intelligence.* Palo Alto: Pacific Books, 1950.
18. Rapaport, D. *Diagnostic psychological testing: the theory, statistical evaluation, and diagnostic application of a battery of tests.* Vol. II. Chicago: Yearbook Publishers, 1946.
19. Rapaport, D. On the psychoanalytic theory of thinking. *Int. J. Psychoanal.*, 1950, **31**, 161–170.
20. Rorschach, H. *Psychodiagnostics.* New York: Grune & Stratton, 1942.
21. Schachtel, E. Projection and its relation to character attitudes and creativity in the kinesthetic responses. *Psychiatry*, 1950, **13**, 69–100.
22. Singer, J. L. Delayed gratification and ego-development: implications for clinical and experimental research. *J. consult. Psychol.*, 1955, **19**, 259–266.
23. Singer, J. L., & Spohn, H. E. Some behavioral correlates of Rorschach's experience-type. *J. consult. Psychol.*, 1954, **18**, 1–9.
24. Singer, J. L., & Herman, J. Motor and fantasy correlates of Rorschach human movement responses. *J. consult. Psychol.*, 1954, **18**, 325–331.
25. Singer, J. L., & Opler, M. K. Contrasting patterns of fantasy and motility in Irish and Italian schizophrenics. *J. abnorm. soc. Psychol.*, 1956, **53**, 42–47.
26. Small, K. Planning as a non-intellective component of intelligent behavior. Unpublished doctor's dissertation, Teachers Coll., Columbia Univer., 1954.
27. Thurstone, L. L. *Multiple-factor analysis.* Chicago: Univer. of Chicago Press, 1949.
28. Thurstone, L. L. The dimensions of temperament. *Psychometrika*, 1951, **16**, 11–20.
29. Weisskopf, Edith A. A transcendence index as a proposed measure in the TAT. *J. Psychol.*, 1945, **19**, 317–327.
30. Werner, H. Motion and motion perception: a study of vicarious functioning. *J. Psychol.*, 1945, **19**, 317–327.
31. Werner, H., & Wapner, S. Sensory-tonic field theory of perception. *J. Pers.*, 1949, **18**, 88–107.
32. Werner, H., & Wapner, S. Toward a general theory of perception. *Psychol. Rev.*, 1952, **59**, 324–333.
33. Wilensky, H. The performance of schizophrenic and normal individuals following frustration. *Psychol. Monogr.*, 1952, **66**, No. 12 (Whole No. 344).
34. Wittenborn, J. R. A factor analysis of Rorschach scoring categories. *J. consult. Psychol.*, 1950, **14**, 261–267.
35. Zubin, J. Objective evaluation of personality tests. *Amer. J. Psychiat.*, 1951, **107**, 569–576.

Journal of Consulting Psychology
Vol. 20, No. 5, 1956

The Development of a Short Form of the Manifest Anxiety Scale[1]

A. W. Bendig
University of Pittsburgh

Two studies (1, 2) have indicated that a majority of the 50 items on Taylor's Manifest Anxiety Scale (MAS) are not valid in predicting clinical criteria of manifest anxiety. The consistency in item validity results of independent studies suggests that a shortened form of the MAS, retaining only the valid items, might be more useful and clinically valid than the standard MAS. However, such a shortened MAS might lack the desirable normative characteristics, such as high reliability, of the longer MAS.

The 20 most consistently valid MAS items in the reports noted above (Item 1–16 and 29–32 in 2, p. 360) were selected as a shortened form of the MAS. The 50-item MAS was administered to 744 college *Ss* and their papers scored for the 20-item scale. The 20-item scale, omitting the 30 "nonvalid" items, was given to an additional 324 *Ss*. No significant differences in scale means or variances were found for the two methods of administration, nor were any sex differences evident. A survey of studies using the 50-item MAS shows its median internal consistency reliability to be .82, while the similar reliability of the 20-item scale was .76. Three scores were obtained for 100 random *Ss* who had taken the 50-item form: (*a*) score on all 50 items, (*b*) score on the 20 "valid" items, and (*c*) their scores on the 30 "nonvalid" items. The reliabilities of the three scores were: (*a*) .78, (*b*) .76, and (*c*) .48. The intercorrelations among the scores were: (*a*) and (*b*), .93; (*a*) and (*c*) .91; and (*b*) and (*c*), .70.

It was concluded that the 20-item Pittsburgh revision of the MAS (*a*) has eliminated from the standard MAS items of low internal consistency and validity; (*b*) provides scores that are about as reliable as the 50-item MAS and are highly related to scores on the standard form; and (*c*) is more parsimonious of testing time and probably more valid than the longer MAS.

Brief Report.
Received June 25, 1956.

[1] An extended report of this study may be obtained without charge from A. W. Bendig, Dept. of Psychology, University of Pittsburgh, Pittsburgh 13, Pa., or for a fee from the American Documentation Institute. Order Document No. 4990 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

References

1. Buss, A. H. A follow-up item analysis of the Taylor Anxiety Scale. *J. clin. Psychol.*, 1955, 11, 409–410.
2. Hoyt, D. P., & Magoon, T. M. A validation study of the Taylor Manifest Anxiety Scale. *J. clin. Psychol.*, 1954, 10, 357–361.

Journal of Consulting Psychology
Vol. 20, No. 5, 1956

Perception of Apparent Motion in Organics and Schizophrenics

Rayford T. Saucer
VA Hospital, Perry Point, Maryland
and Herdis L. Deabler
Gulfport Division, VA Center, Biloxi, Mississippi

Klebanoff, Singer, and Wilensky, in their article summarizing (3) recent findings concerning the psychological consequences of brain lesions and ablations, make the following statement:

> Perceptual research with brain-injured patients suggests that the repertory of the clinical psychologist in the examination of neurological patients may soon be augmented by the use of specialized laboratory methods such as tachistoscopic presentation, critical flicker fusion frequency testing, apparent movement thresholds, and other so-called "brass instrument" techniques. The value of these perceptual techniques in localization of lesions remains equivocal. It seems clear, nevertheless, that subtle perceptual impairments or reorganizations following cerebral injury are most clearly brought out by intensive laboratory examination rather than by the widely used global clinical methods . . . (3, p. 11).

The present paper is a report of an experiment in the perception of apparent motion, together with its theoretical bases and implications, and attempts to show that as a result of increased knowledge in this area, future research in it may lead to the development of a reasonably critical diagnostic instrument.

Werner and Thuma (10) found that brain-damaged children failed to report perception of tachistoscopic apparent motion. Relating the deficit to Goldstein's theory of perception, they attributed it to inability to organize figure-ground relationships, which they held to be a fundamental principle of neurophysiological organization. Shapiro (9) attributed the deficit to satiation in the cortex, equating the effect of continuous visual stimulation to the effect of organic damage, in accordance with Köhler and Wallach's (4) theory of satiation.

Price and Deabler (6), in a study designed to extend Freeman and Josey's (1) work on perception of the apparently moving afterimage induced by a rotating Archimedes spiral, report that 98 per cent of the known organic patients tested had difficulty in perceiving or failed to report this distinctive illusion. In discussing their results, they suggest that either a field theory or a theory of perception proposed by one of the writers may be used to explain their findings.

Earlier work in the field of apparent motion by one of the writers (7, 8) has led to the development of new techniques for the study of apparent motion. In the present study, apparent motion is produced by use of two alternately flashing light bulbs which are $1\frac{1}{4}$-inch neon-filled, fluorescent coated Westinghouse Nite-Lites mounted on a vertical black board. These are illuminated alternately, with zero pause time between flashes, by a multivibrator square wave generator, the circuit of which is conventional.

It is found that varying the rate of alternation of light flashes from 2 to 10 per second causes the light stimuli to appear in three phases which are directly related to the frequency of flashes. As the frequency increases, the stimuli may be perceived and described as (*a*) successive alternating appearance of each light; (*b*) apparent movement wherein the two lights seem to fuse into one light which appears to swing smoothly from side

to side; and (*c*) simultaneous appearance wherein the two lights are seen and perceived as flickering. The distinctive apparent movement phenomenon perceived in Phase *b* was observed by Wertheimer in 1912; he referred to it then as "optimal movement" (11). It has since been investigated by numerous psychologists (2, p. 899), being uniformly designated as Beta movement, by which term it will be referred to in this study.

If the light stimuli are made diffuse, with no sharply defined boundaries, such as is accomplished by use of a ground-glass screen placed about one inch in front of the bulbs, then the last state, Phase *c*, may result in a phenomenon earlier reported (7) called Omega movement. Omega movement appears as movement of a dark, quasi-formless, amorphous object. It has movement but differs from Beta in that it does not have the attributes of the primary stimulus. It has a minimal amount of "thing" quality and is frequently reported as "a dark shadow moving back and forth."

The body of evidence indicates a deficit in organics in their ability to perceive apparent motion. It is felt that the deficit reported by Price and Deabler may be a phase of a general disturbance of the perception of apparent motion. The present study is designed to explore further the nature of this deficit.

A theory of motion perception has been previously presented (7, 8). This theory states that motion perception is based on a matrix synthesis of perceptual organizations isomorphic or related to stimulus detail and to temporal aspects of the stimulus configuration. Beta motion, in which the apparently moving object has the attributes of the primary stimulus, may be regarded as a synthesis of stimulus detail and motion. When the rate of alternation is increased from that which gives optimal Beta motion, the complex or synthesis breaks apart or dissociates. Perception of Beta movement is lost and the observer reports seeing either the two lights as stationary or the Omega movement phenomenon. Practically speaking, the threshold for loss of Beta is the same as for appearance of Omega. It is felt that the ability to sustain Beta perception at higher levels of complexity is dependent upon the functional efficiency of the cerebral cortex. The threshold for loss of Beta perception may thus lend itself as a measuring index for determining cortical efficiency or impairment.

It is to be expected that organics with their decreased cortical efficiency will have lower thresholds for loss of Beta or for appearance of Omega than normals. Schizophrenics by operational definition and by their performance on most psychological tests are known to have low cortical efficiency also. They are included in this study to determine whether or not this is reflected in lower Omega threshold or whether this psychiatric group remains essentially normal with respect to ability to perceive apparent motion.

It is the specific hypothesis of this study that as the rate of alternation of the two lights is changed, organics will have a lower experimental threshold for Omega or loss of Beta than normals, and it is highly probable that schizophrenics will also be lower than normals.

Subjects and Method

Four groups of 10 male subjects per group were used in this experiment, three of the groups being drawn from the patient population of a psychiatric hospital and one group from the population of nursing aides and attendants. The groups were designated as organic, schizophrenic, lobotomized, and control.

Diagnosis of patients by the hospital staff was used as the criterion for grouping. The official diagnosis for each patient represented the opinion of psychiatrists, psychologists, and occasionally, psychiatric consultants. Patients were selected from the current case book of the Psychology Service and represented a random sampling of a hospital population. It is believed that they were typical for their particular diagnostic groups.

Although many organic subjects had a history of one or more brain injuries, no particular category of organic involvement was favored. Alcoholic and senile syndromes were excluded. Practically all organic patients tested carried a secondary or qualifying diagnosis of "with psychotic reaction."

Schizophrenics were taken from the category of schizophrenic reaction, undifferentiated type. If a history of concussion damage, blackouts, or other disorders of consciousness was elicited, subjects were not used. All schizophrenic patients used in the experiment had experienced delusions or hallucinations at some period during their psychosis. Patients with

a history of ECT or ICT during the previous three months were not used.

Ten Freeman-Watts lobotomy cases were also included in the study for comparative purposes. All were diagnosed chronic schizophrenics prior to the lobotomy.

The controls were selected at random from male aides employed in the psychiatric hospital. They were used because it was felt that they represented a group of individuals of average intelligence and similar cultural background to the majority of the patients used.

Subjects were brought into a darkened room and seated two meters from the light source. In a pilot study, it was found that a flash rate of 2.5 flashes per second gave optimal Beta movement for most subjects. This rate was then used as a baseline for the experiment. The rate control was set at 2.5 f.p.s., and each subject was asked to report what he saw. Of the 40 subjects tested, two organic subjects failed to report Beta movement; of these, one reported Omega and another reported seeing no motion of any description. The flash rate was then raised in steps of 0.5 f.p.s. every 30 seconds until the subject reported Omega movement. In a few cases the Beta organization was broken up by simultaneous appearance of two fused lights, which represented a high threshold of dissociation. This point was equated to the perception of Omega motion (8).

Results

The rate at which Omega motion became visible or the rate at which the two lights appeared simultaneously was taken as the threshold of dissociation of Beta movement. The mean rates for the diagnostic groups and their associated standard deviations are given in Table 1.

It can be seen from Table 1 that the means can be ranked from organic through normal. Analysis of the significance of the difference of the means is contingent upon the results of the analysis of variance and will be presented in that order.

Table 1

Threshold Means and Standard Deviations of Groups

Measure	Organics	Schizophrenics	Lobotomies	Control
Means	3.58	3.96	4.45	4.52
Sigmas	.22	.13	.22	.14

Table 2

Analysis of Variance

Source	Sum of Squares	*df*	*F*
Total	19.19	39	
Between	5.94	3	5.50
Within	13.26	36	

Since a test for homogeneity of variance indicated that the variance of the groups represents no more than chance distribution, analysis of variance for between and within categories was applied to the data. The results are given in Table 2.

For 3 and 36 degrees of freedom, an F ratio of 5.50 is significant beyond the .01 level but not to the .001 level of significance. The null hypothesis, that these groups represent random samples drawn from a common population, can be rejected at a satisfactory level of confidence. It may then be concluded that the threshold for the dissociation of Beta apparent motion is related to the presence of pathology.

Following McNemar (5, pp. 74–75), the standard error of the difference can be computed by using the within-variance as the best estimate of the population variance. If this is done, the standard error of the difference for the groups can be based on $N - k$ degrees of freedom, where N is 40 and k is 4. Thus, 36 degrees of freedom are available for the various tests. The various tests are summarized in Table 3.

Table 3

Significance of Mean Differences

Groups Compared	*t*	Significance
Organic-Schizophrenic	1.5	NS
Organic-Lobotomy	3.3	.01
Organic-Normal	3.5	.01
Schizophrenic-Lobotomy	1.7	.05
Schizophrenic-Normal	2.1	.025
Lobotomy-Normal		NS
OS-LN	3.78	.01
OS-N	3.40	.01

For four of the pairs, a t which is not quite significant at the .001 level of significance but well beyond the .01 level can be established. Two are not significant and two are significant at the .05 level of confidence. Current statistical practice sets no level for rejecting the null hypothesis, although the writers feel that the confidence level should at least exceed .01 before the null hypothesis is rejected. If we wish to consider these ratios as a *CR* (since 36 degrees of freedom are involved), then those ratios above 3.3 rise considerably above the .001 level of significance. The writers prefer the more conservative evaluation.

Discussion

The primary hypothesis that organics differ from normal or control subjects in Omega threshold has been established at a highly significant level of significance. Thus we may expect that as test procedure and instrumentation are refined, a diagnostic instrument may be evolved which will provide a statement of degree of pathology but not necessarily indicate the kind. Although schizophrenics are not distinguishable from organics on the basis of this technique, the distinction between normals and schizophrenics suggests that an apparent-motion perception technique may assist in differentiating these conditions.

At the same time, significant F ratio indicates a significant nonchance order of mean scores. If the F ratio be taken as a limited case of a correlation coefficient, the Omega threshold is then correlated with some unknown function. This function is not necessarily the degree of behavioral deviation exhibited by the patient. The writers believe that the means represent points on a continuum of cortical efficiency.

If the results are examined in the light of a neural network frame of reference, they become meaningful. A matrix theory of perception can be expanded to include the entire cerebral cortex as a single matrix. Any damage to any part of this matrix represents a loss and hence lowers the functional efficiency of the matrix. One of the major functions of the cortical matrix is to mediate contact between the organism and the environment by ordering the information received through sensory channels. The subjects selected as organic demonstrate damage to the visual perceptive mechanisms regardless of the type of cortical damage. Evidently perception cannot be fractioned into visual or other components but is a global effort of the entire cortex. The continuum on which the subjects are ranked is the continuum of over-all efficiency of the cerebral cortex, not the efficiency of the visual pathways and visual projection areas alone.

The organic subject exhibits deficit because of damage to neural processes. The schizophrenic, on the other hand, shows deficit because of psychological factors which may or may not have a demonstrable isomorphic physiological basis. A large part of the cortex of the schizophrenic is involved in dealing with problems arising from emotions such as hostility toward a basically unfriendly environment. Neural matrices normally used in mediating environmental contact are disrupted by being organized into matrices concerned with problems which we may broadly define as emotional or as internal. Such neural mechanisms as are involved in these autistic processes must be subtracted from the sum-total amount of cortical tissue available to the organism for the mediation of contact with the environment. By a law of mass action, the results are not to be clearly separated from the results of organic damage.

The interpretation of schizophrenic responses may be inferred from the fact that when lobotomized patients are tested they fall toward the "normal" rather than the "deficit" end of the continuum. As has been mentioned, these patients were chronic schizophrenics unimproved by other therapies. When the subcortical tracts which mediate emotional responses to sensory material are out, they tend to return to the norm with respect to perceptual organization. As the tissue damage is mainly in subcortical areas, the functional Gestalt of the cortex is not destroyed.

From the experimental evidence, it may be inferred that perception is a global process of the entire cortex and that by measuring the amount of organizational force available to the individual for use in perceptual processes such as perception of apparent motion, the functional efficiency of the individual

may be measured and presence or absence of pathology determined.

Summary

Experiments reported by previous investigators indicated a deficit in the perception of apparent motion by organic individuals. The hypothesis that the deficit is due to lowered organizational ability has been advanced. The degree of organizational ability of the individual would be reflected in the threshold for dissociation of Beta apparent motion into component parts of detail and Omega movement.

Organics, schizophrenics, lobotomy cases, and controls were examined for threshold of Beta dissociation. Analysis of variance indicates that the null hypothesis—that these four groups are random samples of a common population with respect to the variable measured—can be rejected at beyond the .01 level of significance.

It is concluded that the threshold for loss of Beta apparent motion can differentiate between certain clinical groups, and it is inferred that these differences are due to organic or functional loss of organizing capacity of the cerebral cortex as a whole rendering it less able to synthesize sensory data.

Received November 1, 1955.

References

1. Freeman, E., & Josey, W. E. Quantitative visual index to memory impairment. *Arch. Neurol. Psychiat.*, 1949, **62**, 794–796.
2. Graham, C. H. Visual perception. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 868–920.
3. Klebanoff, S., Singer, J. L., & Wilensky, H. The psychological consequences of brain lesions and ablations. *Psychol. Bull.*, 1954, **51**, 1–41.
4. Köhler, W., & Wallach, H. Figural aftereffects; an investigation of visual processes. *Proc. Amer. phil. Soc.*, 1944, **88**, 269–357.
5. McNemar, Q. *Psychological statistics*. New York: Wiley, 1949.
6. Price, A. C., & Deabler, H. L. Diagnosis of organicity by means of spiral aftereffect. *J. consult. Psychol.*, 1955, **19**, 299–302.
7. Saucer, R. T. Nature of perceptual process. *Science*, 1953, **117**, 556–558.
8. Saucer, R. T. Processes of motion perception. *Science*, 1954, **120**, 806.
9. Shapiro, M. B. A preliminary investigation of the effects of continuous stimulation on the perception of "apparent motion." *Brit. J. Psychol.*, 1954, **45**, 58–67.
10. Werner, H., & Thuma, A. A deficiency in the perception of apparent motion in children with brain injury. *Amer. J. Psychol.*, 1942, **55**, 58–67.
11. Wertheimer, M. Experimentelle Studien über das Sehen von Bewegung. *Z. Psychol.*, 1912, **61**, 161–165.

Journal of Consulting Psychology
Vol. 20, No. 5, 1956

Emotional and Intellectual Concomitants of Advanced Chronic Alcoholism

George A. Peters, Jr.[1, 2]
Boston Regional Office, Veterans Administration

Much has been written concerning the effects of excessive drinking on human mental processes. The chronic alcoholic has been variously labeled as psychotic, psychoneurotic, psychopathic, and intellectually deteriorated. However, these beliefs have been based upon studies which have dealt with obviously biased samples drawn from such sources as alcoholics' courts, mental hospitals, psychoanalysts' case loads, and the prisons.

A sample thought to be more representative of chronic alcoholics in the general population was drawn from inpatients on the alcoholism ward of a large Eastern general hospital. A group of 120 males, virtually all the chronic alcoholics admitted for treatment during the prescribed period of this study, were examined. The source of patients was quite varied and there were many case histories of delerium tremens and visual and auditory hallucinosis associated with the use of alcohol. These patients had been drinking excessively for from 10 to 30 years, were acutely alcoholic shortly before admission, and were physically debilitated as a result of prolonged alcoholic excess.

Administration of the Wechsler-Bellevue Intelligence Scale revealed a wide range of mental ability, little variation between the verbal and performance subtest scores, and a slightly higher mean score than found in the general population standardization data. (Full Scale mean 106.2, *SD* 13.89; Verbal mean 106.1, *SD* 14.21; Performance mean 105.1, *SD* 13.72.) Academic accomplishments ranged from below the level of basic education to possession of a doctoral degree. The median school grade completed was the tenth grade.

Diagnosis in accordance with the psychobiologic nomenclature of the American Psychiatric Association indicated that these chronic alcoholics were not primarily psychotic (8.3%), psychoneurotic (7.5%), nor sociopathic (5.0%). The number classified as chronic or acute brain syndrome (5.8%) indicates that only a small minority developed clinical symptoms of central nervous system "organicity." Over half (52.5%) were classified as manifesting symptoms indicative of a Personality Disorder.

This group of patients suffering from the physical complications of advanced chronic alcoholism displayed *no* evidence of marked intellectual deterioration on the basis of gross IQ scores. The majority of cases were best described as manifesting a variety of personality disorders rather than psychoses, neuroses, sociopathic, or brain syndromes. If this sample is used as a criterion, there is a surprising lack of major psychiatric disorder or restriction of intellectual capability in the advanced stages of chronic alcoholism.

Brief Report.
Received April 2, 1956.

[1] Formerly Senior Psychologist, Alcoholism Study Unit, Philadelphia General Hospital, Philadelphia, Pa. Gratitude is expressed for the aid given by Drs. Martin D. Kissen, Thomas Downs, and Albert Biele of the Philadelphia General Hospital.

[2] An extended report of this study may be obtained from George A. Peters, Jr., 65 Manor Drive, Newark 6, New Jersey or for a fee from the American Documentation Institute. Order Document No. 4971 from the ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.75 for microfilm or $2.50 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

Journal of Consulting Psychology
Vol. 20, No. 5, 1956

Repression and Insight as Related to Reaction to Cartoons

Seymour Epstein and Richard Smith
University of Massachusetts

A commonly accepted theory of humor is the Freudian view that repression favors an appreciation of humor in which the repressed material is represented (7). Although several studies on the nature of the humor response have been reported, most of which have endorsed Freud's view, a critical appraisal of these studies indicates that they fail to deal directly with the Freudian hypothesis. What they do reveal is that (*a*) individuals show a consistent preference for certain types of humor (1, 3, 4, 5, 8, 12, 13), (*b*) preference for certain types of humor can be related to other personality variables (2, 4, 5, 9), and (*c*) preference for certain types of humor can be related to adjustment (12).

In order to evaluate the Freudian hypothesis it would be desirable to obtain a measure of preference for a certain type of humor, to secure an independent measure of repression, and to determine whether there is a significant relationship between the two. This is what the present study proposes to do.

Method

Hostility was selected as the variable to be investigated because, as a socially disapproved trait, it is subject to repression and there is a considerable amount of humor devoted to it. In order to obtain a measure of preference for hostile humor, 16 hostile and 16 nonhostile cartoons were obtained by selecting only those unanimously agreed upon by independent decisions from five judges. The cartoons finally selected were assigned random numbers and presented to each *S* individually with two sets of instructions. First *S* was to rate each cartoon by placing it into one of three bins according to whether it was "funny," "mediocre," or "not funny." As *S* observed each cartoon, *E* (examiner) unobtrusively recorded on a checklist whether *S*'s expressive reaction was: (*a*) a negative verbal expression, (*b*) a negative facial expression, (*c*) no reaction, (*d*) a positive facial expression, (*e*) a positive verbal expression. After *S* had completed sorting the cartoons, *E* shuffled them and instructed *S* to rearrange them in a Q sort by placing the funniest cartoons farthest to the right and the least funny farthest to the left. The Q sort contained seven columns with successive frequencies 2, 4, 6, 8, 6, 4, 2. The Q sort differed from the initial ratings in that it required that each cartoon be judged relative to all other cartoons.

Establishing a measure of repression was more difficult than measuring reaction to humor. According to Freud, "The essence of repression lies simply in the function of rejecting and keeping something out of consciousness" (7, p. 86). Accordingly, it would seem reasonable to assume that repressed impulses would achieve expression in everyday behavior to a greater extent than they would be consciously acknowledged. Consequently, an operational measure of repressed hostility was established as the degree to which a person denied the hostility attributed to him by a consensus of persons who knew him over an extended period. A measure of insight was obtained by considering the absolute discrepancy without regard to direction (2, 11).

Thirty-two fraternity members who had known each other for at least a year served as *S*s. A week after having sorted the cartoons, each *S* was again seen individually and presented with 32 cards containing the names of all *S*s, including his own. It was requested

that the cards be placed in a Q sort such that the person who most approached the following description would be farthest to the right: "This person is an individual who tends to be inconsiderate, antagonistic, and resentful. When talking about others he is apt to emphasize their faults although he does not like this done to himself. He has little sympathy for others and enjoys putting them on the spot. He makes disparaging remarks about people and argues for the sake of argument. It is important for him to come out on top and he is apt to be a sore loser in competition." This definition was made to be relevant to the manifestations of hostility that were apt to occur in group living. The Q sort consisted of seven columns with successive frequencies of 2, 4, 6, 8, 6, 4, 2.

Results

An index of repression and insight was obtained by subtracting from the self-rating of hostility on the Q sort, the average rating by others. Inspection revealed that there were nine individuals whose self-ratings were higher than their average rating by others. The *S*s, therefore, were divided into three groups of nine each: (*a*) a group that overestimated its hostility and was moderately lacking in insight, with a mean self-minus-others score of 1.03, (*b*) a group that very slightly underestimated its hostility and was relatively insightful, with a mean of -0.36, and (*c*) a group that considerably underestimated its hostility and was most lacking in insight, with a mean of -1.72. For simplicity, the groups will hereafter be referred to as Repression Groups and attention will be turned to insight only when necessary. Differences between the groups were evaluated by analysis of variance of the measures of expressive reaction, cartoon rating, and cartoon Q sort. As there was some danger that the measures of insight and repression were confounded with absolute level of hostility, the groups were compared on mean hostility ratings by a simple between-within analysis of variance. An F of less than one was obtained indicating that any differences in over-all hostility were well within chance expectancy.

In scoring expressive reactions, a weight of one was given to facial expression and a weight of two to verbal expression. An analysis of variance was then performed which allowed for an evaluation of the following: Repression Groups (R), Subjects in same group (S), Favorable vs. unfavorable reaction (F), Hostile vs. control cartoons (C), and the interactions $R \times F$, $R \times C$, $R \times F \times C$, $S \times F$, and $S \times C$. In terms of the hypothesis under investigation, the critical source of variance is $R \times F \times C$, which indicates whether Repression Groups differ in favorableness of response to hostile relative to control cartoons. This interaction did not approach significance, yielding an F of less than one. A second source of variance concerning Repression Groups is the interaction $R \times C$, which indicates whether the groups differ in amount of reaction, irrespective of kind, to hostile relative to control cartoons. Here, too, the F was less than one, suggesting a lack of relationship. The only source of variance relating to Repression Groups which approached significance was R, which indicates total amount of expressive reaction. The F of 2.68 was below the 3.40 required for significance at the .05 level, but did suggest a trend which is borne out by further findings. The order of the means was 9.69 for the group that moderately overestimated its hostility, 4.94 for the group that slightly underestimated its hostility, and 5.11 for the group that considerably underestimated its hostility. Considering that the last two means are almost identical, it appears that the group that overestimated its hostility tended to be more expressive than the other two. Additional findings from the analysis of variance are as follows: (*a*) There were highly reliable individual differences (.001 level) both in amount and favorableness of expression, (*b*) Favorable expression occurred significantly more often than unfavorable expression (.001 level), and (*c*) The hostile cartoons elicited more favorable reactions than the control cartoons (.01 level).

The data from *S*'s ratings of the cartoons were analyzed in the same manner as expressive reactions. Favorable and unfavorable ratings were summated independently, the "mediocre" rating receiving a weight of zero.

Repression Groups were found to differ significantly (.01 level) on total amount of reaction in a manner confirming the nonsignificant tendency noted for expressive behavior. The only other significant source of variance (.001 level) indicated that the hostile cartoons were rated more favorably than the control cartoons.

The data from the Q sort were evaluated by assigning weights of one to seven to the successive columns into which the 32 cartoons had been sorted. Scores consisted of the weight received on each of the 16 hostile cartoons. Analysis of variance revealed that Repression Groups did not differ significantly either in their reaction to the pooled hostile cartoons or in their evaluation of the individual hostile cartoons. The only statistically significant finding (.001 level) was that some of the hostile cartoons were judged funnier than others. Despite the lack of a reliable relationship between repression and reaction to individual cartoons, the data were inspected for leads for further work. It was observed that the cartoon which was ranked first in funniness by the group that considerably underestimated its hostility was ranked sixth by the group that slightly understimated its hostility and ninth by the group that overestimated its hostility. Thus, preference for this cartoon was directly related to denial of hostility. The cartoon was one by Virgil Partch depicting a wife plunging a revolving egg-beater into the back of her husband, and saying, "Stick 'em up!" It was *E*'s impression, confirmed by independent decisions from three other judges, that this cartoon was the most straightforwardly hostile in the sample. Unfortunately, only two similar cartoons of the "practical joke" variety depicting an unprovoked act of hostility by an adult could be found. In both cases a direct relationship was indicated between repression and preference for the cartoon.

In order to determine whether it is necessary to take into account level of funniness of the cartoons, two separate analyses of variance of Q-sort ratings were done, one based on the four least funny, the other on the four most funny hostile and control cartoons as determined by the mean rating of the cartoon. On the least funny cartoons Repression Groups did not differ in preference for hostile relative to control cartoons but did differ in over-all rating of the pooled cartoons (.05 level). The order of the means was 29.33 for the group that considerably underestimated its hostility, 25.11 for the accurate group that very slightly underestimated its hostility, and 26.00 for the group that considerably overestimated its hostility. Thus, the group that evaluated itself most accurately judged the poor cartoons most appropriately, and the reverse holds for the group that evaluated itself most inaccurately. On the most funny cartoon, Repression Groups did not differ significantly in their reaction to the pooled cartoons, but the order of the means for the same three groups was, respectively, 36.44, 39.56, and 37.00, which again suggested that accuracy of self-evaluation was directly related to appropriateness of reaction to the cartoons. Thus, taken together the findings on the least and most funny cartoons indicated a direct relationship between insight and sense of humor.

In order to determine the relationship between the three measures of reaction to hostile cartoons, Pearson product-moment correlations were computed. To do this, the measure of expressive reaction was expressed as a ratio of favorable reactions to hostile cartoons over favorable reactions to all cartoons. The same index was used for cartoon-ratings. For the Q sort, the score was simply the total weight on the hostile cartoons. The following correlations were obtained: expressive reaction and cartoon-rating, .57; expressive reaction and Q sort, .52; and cartoon-rating and Q sort, .79. All correlations are significant beyond the .01 level (t test, $N = 32$). These results indicate a significant degree of reliability which is at least .52 for expressive reaction and .79 for cartoon-rating and for Q sort. That the measure of expressive reaction correlated lowest with the other two may indicate that it has the lowest reliability or that it measured something unique, or both. There is some evidence in support of the latter possibility in that expressive reaction was the only measure on which significant individual differences were found.

Discussion

Repression and humor. The results of this study fail to support the hypothesis that repression of hostility is related to a general preference for hostile cartoons. Negative results are difficult to interpret as there is no way of knowing whether different operational measures or different populations would have produced different findings. In defense of the measure of reaction to cartoons used, adequate reliability for group differentiation was indicated. Moreover, validity was somewhat supported by the finding that reaction to cartoons was meaningfully related to insight. The measure of repression is somewhat more questionable, but it, too, was meaningfully related to another variable in that it was inversely associated with over-all expressiveness. This relationship reveals a degree of reliability beyond chance, although there is no indication of how high it is. However, since the Q technique for measuring repression required each *S* to judge and be judged by all other *S*s relative to each other, an acceptable degree of reliability would be expected, at least approaching the minimum estimate of .79 obtained from the same Q sort for cartoons. Considering that all *F* values on the interaction of repression and preference for hostile cartoons were much less than one, there is no reason to suspect that higher reliability would have altered the findings. That another operational measure of repression might have produced positive findings remains to be demonstrated. Some question might be raised as to whether a group that overestimates its hostility should be considered as low in repression. If only the accurate group and the group that underrated its hostility are compared, the results are equally negative.

One explanation of the negative findings on repression and humor is that there is no *general* relationship between repressed hostility and preference for hostile cartoons. That there might be a relationship between repressed hostility and a preference for certain types of hostile cartoons was suggested by the results on a cartoon of the practical-joke type which depicted an unprovoked act of hostility. However, the evidence here was inclusive and verification is required. A second possibility is that a relationship would be found if suppression (conscious inhibition) or external frustration of an impulse were investigated rather than repression (unconscious inhibition). This, too, remains to be determined, but seems likely in view of common observations of the interest in sexual jokes of sexually frustrated groups such as adolescents (culturally frustrated) and men in the armed services (geographically frustrated), who are not necessarily repressed. A third possibility for investigation is that a relationship between repression and humor becomes manifest only when repression reaches pathological proportions (10).

Insight and humor. Allport states, ". . . the most striking correlate of insight is the *sense of humor*" (2, p. 222). He cites one study in which a correlation of .88 was found between ratings on insight and ratings on sense of humor. In the present study it was found that there is a statistically significant relationship between rating the cartoons correctly and judging one's hostility accurately. Why are insight and humor related? Although there was no way of testing this in the present study, it is hypothesized that the two are related only when humor involves an ability to accept the self as its object. Presumably the person who is able to laugh at his limitations should be low in defensiveness and accordingly should be insightful. It would be interesting to develop a measure of differential appreciation of humor in which the object of some of the jokes or cartoons is self-related and of others is other-related. This measure might then be related to measures of defensiveness and insight. Perhaps a subtle clinical instrument for determining defensiveness and insight would result.

An incidental finding of some interest was that there are individual differences in degree and type of expressive reaction irrespective of the content of the cartoon responded to. This finding indicates that it is necessary to use control cartoons in evaluating differences in expressive reaction to any particular type of cartoon. It further indicates that individual differences in expressive behavior, apart from the nature of the cartoons responded to, is it-

self a promising personality variable for investigation.

Summary

Preference for hostile humor was measured by three techniques, which included observing *S*'s expressive reactions to hostile and control cartoons, having *S* rate the cartoons, and having him sort them on a Q sort of funniness. In order to obtain a measure of repression, the *S*s, members of the same fraternity for at least a year, rated themselves and each other on a Q sort of hostility. The degree to which the self-rating was lower than the average rating by others was taken as an index of repression; the absolute discrepancy between the self-rating and rating by others was taken as an index of insight. Three groups of nine each were selected, a group that considerably overestimated its hostility, an accurate group that very slightly underestimated its hostility, and a group that considerably underestimated its hostility. Following are the major findings:

1. No relationship was found between repression and general preference for hostile cartoons, but there was questionable evidence that repression was related to preference for a practical-joke type of cartoon depicting an unprovoked act of hostility.

2. A significant positive relationship between insight and sense of humor was found.

3. A significant negative relationship between repression and degree of reactivity was found.

4. Highly reliable individual differences in degree and kind of expressive reaction irrespective of the cartoons responded to were found.

Implications for further work on the relationship of humor to repression and insight were discussed.

Received January 9, 1956.

References

1. Andrews, T. G. A factorial analysis of responses to the comic as a study in personality. *J. gen. Psychol.*, 1943, **28**, 209–224.
2. Allport, G. W. *Personality, a psychological interpretation.* New York: Holt, 1937.
3. Cattell, R. B., & Luborsky, L. B. Personality factors in responses to humor. *J. abnorm. soc. Psychol.*, 1947, **42**, 402–421.
4. Eysenck, H. J. The appreciation of humor: an experimental and theoretical study. *Brit. J. Psychol.*, 1942, **32**, 295–309.
5. Eysenck, H. J. Experimental analysis of five tests of "appreciation of humor." *Educ. psychol. Measmt*, 1943, **3**, 191–214.
6. Freud, S. *The basic writings of Sigmund Freud.* New York: Random House, 1938.
7. Freud, S. *Collected papers.* Vol. 4. London: Hogarth, 1950.
8. Luborsky, L. B., & Cattell, R. B. The validation of personality factors in humor. *J. Pers.*, 1947, **15**, 283–291.
9. Murray, H. A., Jr. The psychology of humor. *J. abnorm. soc. Psychol.*, 1934, **29**, 66–81.
10. Redlich, F. C., Levine, J., & Sohler, T. P. A mirth response test; preliminary report on a psychodiagnostic technique utilizing dynamics of humor. *Amer. J. Orthopsychiat.*, 1951, **21**, 717–734.
11. Sears, R. R. Experimental studies of projection. In S. S. Tomkins (Ed.), *Contemporary psychopathology.* Cambridge, Mass.: Harvard Univer. Press, 1943.
12. Strother, G. B., Barnett, Margaret M., & Apostolakos, P. C. The use of cartoons as a projective device. *J. clin. Psychol.*, 1954, **10**, 38–42.
13. Yarnold, J. K., & Berkeley, M. H. An analysis of the Cattell-Luborsky Humor Test into homogeneous scales. *J. abnorm. soc. Psychol.*, 1954, **49**, 543–546.

Journal of Consulting Psychology
Vol. 20, No. 5, 1956

The Meaning and Generality of Perceived Actual-Ideal Discrepancies

Leon H. Levy[1]
Indiana University

Recent studies by Rogers (4) and his associates have revealed that when subjects (*Ss*) are asked to do Q sorts of self-referent items under instructions to sort them so that they describe themselves as they are at present and then again so that they describe their concept of the ideal person, the correlation between these two sorts seems to vary with the presence or absence, or severity, of maladjustment. The size of the correlation between actual- and ideal-self-sorts has been interpreted variously as a measure of self-esteem or self-acceptance, and the finding that persons classified as maladjusted have lower Q correlations than persons not so classified has been taken as evidence for the importance of self-acceptance in adjustment and the value of the self concept as a construct in personality theory. Other investigators have suggested defining personal maladjustment "as any discrepancy between the concept of self and the concept of ideal self" (1, p. 257).

However, upon careful analysis of the procedures and data involved in these experiments it does not appear that these inferences are entirely justified. In this paper we propose to examine the logical justification for these inferences and then to present data supporting an alternate hypothesis.

The task set for the *S* in the typical actual-ideal Q-sort study is one of describing one aspect of his environment as he perceives it in its actual state and again as he perceives its ideal state to be. The resulting correlation may be interpreted as an indication of the perceived discrepancy between the actual and ideal states of the particular aspect of the environment delineated for the *S* by the experimenter's instructions and the descriptive items he has to work with. In the studies referred to above the aspect of the environment involved is labeled "the self" and a set of verbal items deemed by the experimenter as appropriate in describing the self function to delineate that aspect of his total environment for which the *S* is to indicate his perceived actual-ideal discrepancy.

The finding that a certain portion of the variance of the discrepancy measure may be attributable to individual differences does not in and of itself permit any inferences concerning the significance of the particular aspect of the environment used in obtaining this discrepancy measure. That is, until it has been demonstrated that actual-ideal correlations vary with adjustment if and only if they are obtained through the use of the self as the object of description, inferences concerning the importance of the nature of the self concept in personality functioning are not justified. It may be, for example, that the more maladjusted the individual the greater the discrepancy he perceives between the actual and ideal state of affairs over a wide range of areas of his environment. Thus, conceivably, one might discover a significant relationship between Q correlations for actual-ideal descriptions of geographical climate and personality adjustment. Unless one went on to sample other aspects of the *S*'s environment as objects of description he might be led to construct a theory of personality functioning with climatological perception as a central construct.

[1] The author wishes to thank Mr. Richard F. Gundy who ran the *Ss* and performed the statistical computations.

The hypothesis which we propose to test in the present study is that individuals differ in the extent to which they perceive discrepancies between the actual and ideal states of various aspects of their environment and that these differences are fairly constant from one area to another. Thus an individual obtaining a high Q correlation when doing actual-ideal sorts of self-referent items would be expected to obtain a high Q correlation when doing actual-ideal sorts of items descriptive of other aspects of his environment, and similarly with an individual obtaining a low Q correlation between actual-ideal self-sorts.

Procedure

The *Ss* used in this experiment consisted of 21 volunteers (2 females and 19 males) from an introductory psychology course, none of whom were aware of the nature of the experiment at the time of their volunteering for it.

Each *S* was seen individually and was asked to do actual-ideal sorts of 100 self-referent items and of 100 home-town-referent items. The self-referent items were those developed by Butler and Haigh (2).[2] The home-town-referent items were collected in the following manner. Twenty students in the author's personality class were given the following instructions: "(*a*) List five things that you like most about your home town. (*b*) List five things that you like least about your home town. (*c*) List five things which you feel would be true of the ideal home town." From the resulting pool of 300 items, 100 were selected which appeared to have fairly wide applicability and which provided an approximately equal representation of positive and negative statements.

The *Ss* were required to sort the statements for each sort into nine piles with the following distribution of cards per pile: 1, 4, 11, 21, 26, 21, 11, 4, 1. Each *S* was first presented with the self-referent statements and asked to sort them so as to describe himself as he sees himself at present from those least like himself to those most like himself. Following this he was asked to sort the home-town-referent statements so as to describe his own home town. Next he was asked to sort the self-referent statements again so as to describe his ideal person, or the person he would most like to be. And finally, he was asked to sort the home-town-referent statements so as to describe his concept of the ideal home town, the kind of town he would most like to live in. The same order of presentation was used with all *Ss*.

The two Q correlations resulting for each *S* were then correlated over all *Ss* in order to determine the extent to which actual-ideal discrepancies in the two areas sampled were related.

Results and Discussion

The median Q correlation for actual-ideal self-descriptions was .47 with a range from .08 to .82. The median Q correlation for actual-ideal home-town descriptions was .24 with a range from − .28 to .92. The Pearson correlation between the two sets of Q correlations was .70 which is significant at the .01 level. This finding makes the null hypothesis untenable and gives some support to the experimental hypothesis that the Q correlation obtained for an individual on actual-ideal self-descriptions represents a tendency to perceive actual-ideal discrepancies which transcends the particular aspect of the *S*'s environment used as the object of description. Furthermore, in the light of the present data, previous hypotheses concerning actual-ideal discrepancies which treated them as phenomena contingent upon such conditions as self-acceptance, self-esteem, self-satisfaction, etc., appear to be too narrowly conceived.

While the self concept as an important construct may still be salvaged by invoking the processes of projection and rationalization on the part of the *S* to account for generalized perceived actual-ideal discrepancies, particularly in the case of maladjusted individuals, such reasoning must be recognized as *ad hoc*.

In the present study only two aspects of the *Ss*' environment were used as objects of description. Before any definite statement can be made about the generality of perceived actual-ideal discrepancies many more aspects would need to be sampled. It would appear however that research might be profitably di-

[2] The author wishes to thank Dr. Rosalind F. Dymond for making these items available for this study.

rected toward a study of the nature and extent of perceived actual-ideal discrepancies.

Certain rather interesting research problems become opened up by such an approach and it may be of some value to sketch these briefly. One such problem would be: Given a certain number of objects of description, do individuals vary in the consistency of their perceived actual-ideal discrepancy over the range of these objects? What factors might best account for the variance in this consistency? What factors, if any, appear to be a function of the degree of perceived actual-ideal discrepancy in any given area of description? Do these vary from area to area? The theoretical implications of such research are difficult to anticipate but it seems fairly clear that current self-oriented neophenomenological personality theories (3, 5) would not be adequate to account for these data.

Summary

The logical justification for current interpretations of actual-ideal self-description discrepancies was called into question. An experiment was performed in which *Ss* did actual-ideal Q sorts using self-referent items and using home-town-referent items. The resulting sets of correlations were found to have a statistically significant correlation of .70 between them and this was taken as evidence indicating a certain amount of generality to perceived actual-ideal discrepancies and further substantiating the criticism of current interpretations of these discrepancies when limited to the self as the object of description. Implications of these findings for future research were sketched briefly.

Received December 5, 1955.

References

1. Bills, R. E., Vance, E. L., & McLean, O. S. An index of adjustment and values. *J. consult. Psychol.*, 1951, 15, 257–261.
2. Butler, J. M., & Haigh, G. V. Changes in the relation between self-concepts and ideal concepts. In C. R. Rogers & Rosalind F. Dymond (Eds.), *Psychotherapy and personality change.* Chicago: Univer. of Chicago Press, 1954. Pp. 55–75.
3. Rogers, C. R. *Client-centered therapy.* Boston: Houghton Mifflin, 1951.
4. Rogers, C. R., & Dymond, Rosalind F. (Eds.) *Psychotherapy and personality change.* Chicago: Univer. of Chicago Press, 1954.
5. Snygg, D., & Combs, A. W. *Individual behavior.* New York: Harper, 1949.

Journal of Consulting Psychology
Vol. 20, No. 5, 1956

Learning in the Immediate Post-Electroshock Period[1]

George A. Zirkle
Hanover College

The view has been current for some years that the period soon after electroconvulsive shock offers unique opportunities for psychotherapy with mental patients. Though many clinicians have used the period rather extensively for this purpose, others have questioned its value. For instance, Kalinowsky and Hoch write, "In our experience, psychotherapy during convulsive therapy has often been impossible because of the patient's confusion" (3, p. 118).

How valuable is the postshock period for psychotherapeutic purposes? If it is valuable, how soon after shock should psychotherapy be instituted? Answers to these questions will depend, among other things, on an analysis of the mental state of the patient after disorganizing shock, and a determination of the time pattern by which abilities are recovered. A number of research efforts have been directed toward this end, most of which have concentrated on memory functioning. Certain investigators (1, 5, 6) have reported a decided interference of shock with memory. Experimental testing has usually been done several hours or more after shock, rather than immediately afterward. An example of work in the immediate postshock period, however, is that of Lunn and Trolle (4), who studied mental abilities 10, 30, and 60 minutes after shock. They note very poor retention for materials presented shortly after shock. Thirty minutes afterward, however, they report that 85 per cent of their patients were able to give "a satisfactory description of the content of a picture and at the same time express a comprehension of the point of the picture" (4, p. 376).

[1] The cooperation of Ott B. McAtee, M.D., Superintendent, William E. Murray, M.D., Asst. Superintendent, and the staff of Madison State Hospital in making their facilities available for this study is hereby gratefully acknowledged.

The Problem

Up to this time, relatively little research attention has been given to one of the crucial factors in the ability to profit from psychotherapy—namely, *learning*. This is especially true of learning in the immediate postshock period, when psychotherapy is often instituted. As a partial answer to this lack of information, the present research was designed to study, under controlled conditions, the pattern for recovery of learning ability after shock.

One possible, and no doubt partial, interpretation of the therapeutic effect of convulsive shock is that it temporarily disrupts or disorganizes abnormal mental patterns and permits the substitution of different patterns of a more normal or healthy character. We are well aware of the fact that a startling event can shock a normal person "out of his complacency" and cause him to see familiar things in new ways. It may be that convulsive shock does somewhat the same thing for the abnormal person, and jolts him out of his "psychotic rut." If this line of thought be followed, then it would appear that the optimum time to substitute new patterns of thought would be at the point where disorganization is great enough to permit their introduction and yet where learning ability is sufficiently recovered to permit their retention. In other words, if the above interpretation has any validity, the best time for a reeducative therapy would be as soon after shock as a reasonable degree of learning ability has returned.

Subjects. Subjects in this investigation were five female and three male patients at Madi-

son State Hospital who were receiving electric convulsive therapy. They were all literate, in relatively good contact, and cooperative. Their ages ranged from 27 to 62, the average being 37.1. Six were schizophrenic, one manic-depressive, and one involutional.

Experimental materials. Three types of learning materials were developed and used —pictorial, verbal, and numerical. To test visual object learning, a series of 16 line drawings was prepared. Each drawing contained a mannikin surrounded by certain objects. Fifteen details were catalogued in each picture, such as type of hat, tie, etc. on the mannikin, and house, balloon, and other objects surrounding the mannikin. Five different objects or conditions were systematically varied at each of the 15 locations on the different pictures. For instance, the objects on the right side of the mannikin were a bicycle, a clock, a suitcase, a tree, and a truck; and the mouth of the mannikin was shown variously as open, straight, corners up, corners down, and with one corner up and the other down. The subject was shown a picture and each detail was pointed out and verbally called to his attention with the request that he remember as many as possible. Each picture also contained a different mental hygiene aphorism having eight words and five ideas, which was attributed to the mannikin. An example was, "Interests outside make a person far happier inside." Patient motivation for the task was good and several seemed to enjoy it. Indeed, success with this procedure would seem to dictate its wider use as a measure of learning ability.

To test verbal learning, 16 lists of paired adjectives were developed. Each list contained 11 pairs. They were obtained from a list of 300 pairs developed by Melton and Safier (2), and were relatively equivalent in terms of the similarity ratings given them by 96 university students. A list was slowly read over once to a subject with the instruction that he should try to remember as many of the pairs as possible. Thereupon he was handed a typewritten list containing the first adjective in each pair ranged on the left margin, and on its right the second member imbedded among three other synonymous adjectives. For instance, the pair *Pleasing-Charming*, appeared on the typed list given to him as follows: *Pleasing: Pleasant, Gracious, Charming, Winning*. The subject was asked, "Which word did I read with *Pleasing?*" The same procedure was employed with each of the other pairs.

To investigate numerical learning, a series of 16 lists of numbers was prepared. Each list contained two samples each of two-digit through eight-digit numbers, arranged as in the Wechsler-Bellevue Intelligence Scale. Starting with the shorter numbers, subjects were asked to repeat each one immediately after it was read to him. His score was the highest number he could accurately repeat. No attempt was made to go further when both samples of a given length were failed.

Procedure. Subjects were enlisted by explaining to them that they had been chosen for a series of memory tests. One practice trial was given on each of the three types of materials to familiarize the subject with the procedure. The order of presentation was picture, paired adjectives, and numbers.

A control series of tests was run with each subject one to three days before shock was begun. Runs of the three types of tests were given six different times. Some hours to a day intervened between the first three and the last three runs. A short break of a minute or so was introduced between runs.

Experimental testing was begun as soon after shock as the patient was able to get up and go to a nearby examining room. Two runs of tests were given in succession. A third run was given after the patient had had lunch. This procedure was followed after three shocks—in practically all cases the first three. This made a total of 9 experimental runs as compared to 6 control runs. Runs were rotated so that four subjects got the test series in the order 1 to 15, and the remaining four in the reverse order.

The time for beginning of experimental testing after shock varied from 10 to 30 minutes, with an average for all subjects of 18½ minutes. Several patients became nauseated, showed confusional retardation, or displayed paranoid resistance. All these caused testing delay.

Results

Table 1 shows clearly the effect of shock on learning performance. A profound interference with learning ability is shown in the first test run after shock, which was an average of 18½ minutes later. The critical ratios of differences between *experimental* and *total control* means are highly significant except in

Table 1

Scores on Control and Experimental Series

Test and measure	Control runs				Experimental runs		
	1	2	3	Total	1	2	3
15 Mannikin Items							
Mean	10.8	11.3	11.4	11.2	2.5	8.3	12.2
Percentage of success	72	75	76	75	17	55	81
SD				2.21	2.96	2.87	2.21
CR with control total					12.39	4.23	1.82
5-Point Mannikin Aphorisms							
Mean	3.6	2.7	3.2	3.2	.4	1.8	3.4
Percentage of success	72	54	64	64	8	36	68
SD				1.87	1.26	1.87	1.73
CR with control total					7.26	2.89	—
11 Paired Adjectives							
Mean	7.6	7.3	8.6	7.8	4.5	5.5	7.7
Percentage of success	69	66	78	71	41	50	70
SD				2.47	1.66	2.22	2.03
CR with control total					6.76	3.99	—
8 Numbers							
Mean	6.1	6.1	6.0	6.0	5.9	6.0	6.5
Percentage of success	76	76	75	75	74	75	81
SD				.93	1.38	1.23	1.04
CR with control total					—	—	1.26

the case of numerical learning. Interference is noted in less but still significant degree on the second experimental run, except in the case of numerical learning. The second testing run began an average of 38⅔ minutes later. Learning performance has returned to preshock levels by the third testing run, which was an average of 2⅔ hours later.

Discussion of Results

It is clear from these results that electric shock profoundly interferes with learning abilities. Greater interference is shown for abstract than for more concrete materials. The most extensive interference occurred in the case of the aphorisms, which were the most abstract materials used in this experiment. During the first test run after shock, only two persons of the eight could remember any part of the aphorisms. One of these two could not be tested until 30 minutes after shock. The least interference occurred in the case of numerical learning and report. This task represented the lowest degree of abstraction. Hetherington (1) also found little influence of shock on immediate number recall.

Learning was measured in this experiment by means of immediate report. No attempt was made to determine whether materials were retained over a longer period of time. In a re-educative therapy program, long-term retention would be necessary, of course.

So far as learning ability is concerned, it would appear highly unprofitable to attempt a re-educative type of therapy immediately after shock. This is deduced from individual as well as group scores. While individual variations in recovery rate were smoothed by combining them to produce a group picture, all individual patterns of recovery closely paralleled the group pattern.

By two to three hours after shock, results show that learning ability has returned to preshock levels. It is quite probable that control level ability may return even earlier. Timing for the third test run was set at the completion of the lunch hour for patients and so this was not determined. The shortest period for beginning the testing of any patient on the third run of tests was 1 hour and

55 minutes after shock. In this one case learning ability was back to preshock level.

Insofar as present results are concerned, re-educative therapy is contraindicated for most patients during the first hour after shock. Recovery rates varied for different patients, but preshock learning performance is recovered somewhere between 1 hour and 2½ hours after shock. According to the theory enunciated earlier in this article, somewhere within this interval would seem to be the optimum time for the institution of re-educative efforts.

Summary

Using pictorial, verbal, and numerical materials, the learning ability of eight mental patients was tested both before and at stated intervals immediately after administration of electroconvulsive shock. Results showed a profound interference of shock with learning ability. Interference was greater for abstract than for more concrete learning materials. Learning abilities were recovered in a gradual fashion, with preshock level being restored not later than two to three hours after shock. Implications of the results for a re-educative type of therapy are noted.

Received January 3, 1956.

References

1. Hetherington, R. Psychological changes due to E.C.T. as measured by a battery of tests given repeatedly. *J. ment. Sci.*, 1952, **98**, 334–341.
2. Hilgard, E. R. Methods and procedures in the study of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 517–567.
3. Kalinowsky, L. B., & Hoch, P. H. *Shock treatments, psychosurgery*. (2nd Ed.) New York: Grune & Stratton, 1952.
4. Lewis, N. D. C. Psychiatry. In R. P. Mackay, N. D. C. Lewis, & P. Bailey (Eds.), *The 1949 year book of neurology, psychiatry, and neurosurgery*. Chicago: Year Book Publishers, 1950. Pp. 209–423.
5. Stone, C. P. Losses and gains in cognitive functions as related to electro-convulsive shocks. *J. abnorm. soc. Psychol.*, 1947, **42**, 206–214.
6. Zubin, J. Objective studies of disordered persons. In T. G. Andrews (Ed.), *Methods of psychology*. New York: Wiley, 1948.

Journal of Consulting Psychology
Vol. 20, No. 5, 1956

Note on "Changes in Psychoneurotic Patients With and Without Psychotherapy"

Desmond S. Cartwright
University of Chicago

A recent paper in this Journal by Barron and Leary (2) reports a comparison of changes in a treatment group of psychoneurotic patients with changes in a nontreatment group. Groups were matched for type and severity of disorder, prognosis, socioeconomic variables, and interval of time between pretests and posttests.

Of the patients in treatment, 42 received individual and 85 received group psychotherapy. The measures employed were MMPI scales, including a prognostic scale (*Es*) developed by Barron (1). The crucial data showed mean differences in scale scores between pretests and posttests. Comparison of these data by critical ratio analysis led to the conclusion that: "The therapy patients did not improve significantly more than did the waiting-list controls" (2, p. 245).

This conclusion answered the question: ". . . is there significantly more change with regular and formally defined psychotherapy than without it?" (2, p. 245). Statistically, the answer given states that the treatment and nontreatment samples must be considered to be drawn from the same parent population in regard to mean changes. The presence or absence of psychotherapy makes no difference.

However, from the data presented (2, p. 243), it appears that the samples differ as to the dispersion of changes (*SD*s). Table 1 presents the relevant data for the individual psychotherapy group and the nontreatment group, together with *F* ratios computed from the variances.

Table 1 suggests with high probability that the two samples are not drawn from the same parent population in regard to the variance of changes on scales *F*, *Hs*, *D*, *Pt*, *Sc*, and *Es*. This means that the presence or absence of individual psychotherapy makes a considerable difference.

The implication of the results in Table 1 is that in fact there was "significantly more change with regular and formally defined psychotherapy than without it."

Table 1

Variances of Discrepancy Scores on MMPI Scales for Individual Psychotherapy and Nontreatment Groups

MMPI Scale	Individual psychotherapy ($N=42$) V*	Nontreatment group ($N=23$) V	*F*	*p*
1. Lie	19.89	23.43	1.18	—
2. *F*	215.21	22.94	9.38	.01
3. *K*	55.95	31.70	1.76	—
4. *Hs*	127.46	64.16	1.99	.05
5. *D*	244.30	93.32	2.62	.01
6. *Hy*	113.21	87.80	1.29	—
7. *Pd*	155.00	89.68	1.73	—
8. *Pa*	111.94	68.06	1.64	—
9. *Pt*	208.51	73.27	2.85	.01
10. *Sc*	272.91	74.13	3.68	.01
11. *Ma*	126.79	75.34	1.68	—
12. *Es*	43.56	14.82	2.94	.01

* Variances computed from *SD* data reported by Barron and Leary (2, p. 243).

The greater amount of change, however, was not all in one direction. Thus the authors' conclusion that "The therapy patients did not improve significantly more than did the waiting-list controls" is correct but insufficient. For many scales the variance results suggest that mean differences between

the groups are absent because differences of two kinds, opposed in sign, are present. It seems that some therapy patients *deteriorated* to a greater extent than did the waiting-list controls, while some therapy patients *did improve* significantly more than the controls.

Received November 9, 1955.

References

1. Barron, F. An ego-strength scale which predicts response to psychotherapy. *J. consult. Psychol.*, 1953, **17**, 327–333.
2. Barron, F., & Leary, T. F. Changes in psychoneurotic patients with and without psychotherapy. *J. consult. Psychol.*, 1955, **19**, 239–245.

Journal of Consulting Psychology
Vol. 20, No. 5, 1956

A Short Form of the Wechsler-Bellevue and the Army Classification Battery as Measures of Intelligence[1]

Donald B. Murphy
Army Medical Service School, Ft. Sam Houston, Texas

and Robert D. Langston
Br. U. S. Disciplinary Barracks, Ft. Crowder, Mo.

Although the Army Classification Battery and short forms of the Wechsler-Bellevue have been used as measures of intelligence, a review of the literature indicated the application of previously published short forms to an Army disciplinary barracks population was somewhat questionable. The present study was designed to investigate the validity of a short form of the W-B and of Aptitude Area I (AAI) of the Army Classification Battery as measures of intelligence for this population.

For the present study, the records of 515 consecutive releasees from the Br. U. S. Disciplinary Barracks, Ft. Crowder, Missouri, were reviewed. The ages of this group of male prisoners ranged from 16 to 41 with the mode falling at the 20–24 age grouping. Eighty-five per cent worked at unskilled labor or had no occupation prior to induction. Negroes comprised 30% of the group.

The W-B full scale IQ, short form IQ based on the Information, Comprehension, and Block Design subtests, and the scores on the AAI, which has a normative population mean of 100 and *SD* of 20, were obtained for each case.

[1] An extended report of this study may be obtained without charge from Capt. Donald B. Murphy, AMSS, Ft. Sam Houston, Tex., or for a fee from the American Documentation Institute. To obtain it from the latter source, order Document No. 4970 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

The mean of the full scale W-B was 93.2 with an *SD* of 16.1; the mean of the short form was 92.8 with an *SD* of 16.9; and the mean of the AAI was 81.7 with an *SD* of 22.2. The short form of the W-B and the AAI correlated .91 and .83, respectively, with the full scale W-B.

The advantage of the short form, since it is a reliable predictor of full scale scores, is that testing time is reduced, permitting rapid processing and more time for evaluation of nonintellectual factors.

The *Ss* used in this study had a significantly lower IQ ($p < .01$) than Wechsler's normative group and differed from Wechsler's population with respect to occupational status, education, and race. It is interesting to note that the short form used in this study has been reported to have the highest correlation with the full scale W-B for Wechsler's normative group, thus indicating the stability of these subtests despite differences in socioeconomic, educational, and racial backgrounds.

The mean for the AAI was significantly lower ($p < .01$) than the AAI mean (90.7) estimated on the basis of the *Ss*' W-B scores. Marked increase in the variability of AAI scores below 75 suggested that many of the low scores resulted from poor motivation at the time the test was administered rather than from low intelligence. When AAI scores below 75 were discarded, the mean of the AAI scores (93.4) closely approximated the estimated AAI mean and correlated .88 with the full scale W-B.

Brief Report.
Received April 3, 1956.

Journal of Consulting Psychology
Vol. 20, No. 5, 1956

"Reliability of the Blacky Test": A Reply to Charen

Gerald S. Blum
University of Michigan

After reading Charen's brief report on the retesting of 25 tuberculous patients, printed in the February issue of this Journal (5), I was left with two major misgivings: (*a*) the title is a misnomer inasmuch as the technique was used improperly; and (*b*) the author's reasoning is faulty. I shall elaborate these points below:

1. The Blacky Pictures are designed to yield conflict scores for a series of psychosexual dimensions. As stated in the Manual (3), four separate sources contribute to each dimensional score: spontaneous story, inquiry items, related comments, and cartoon preferences. The underlying conception of reliability, upon which the instrument was constructed, assumes that these sources tap the conflict area at varying levels and in combination provide the best possible estimate of conflict intensity. The revised scoring system (4) referred to by Charen makes clear this necessity for including all four sources in deriving a reliable dimensional score. Instead, Charen arbitrarily singled out for investigation one source, multiple-choice items from the inquiry, which is stated to be closest to consciousness and, therefore, most easily distorted by the subject. Also the technique was administered incorrectly, since spontaneous stories were not obtained prior to the responses to multiple-choice items.

2. From the facts that the Rorschach and paper-and-pencil tests indicated no "basic changes" in personality whereas the Blacky items did show some (basic?) changes, Charen infers only two possible alternatives: "Either the Blacky questions tapped personality levels not reached by the other tests or did not have good reliability." He rejects the first alternative by a statement of faith in the Rorschach. I prefer to call upon evidence contained in a study by Berlow (2), who compared Blacky scores (in a sample of patients whose responses had previously been meaningfully related to theories of paranoid schizophrenia [1]) to psychosexual indicators on the Rorschach. In addition to using judges to assess psychosexuality from the Rorschach, Berlow conducted a systematic and exhaustive analysis of signs, patterns, and content. No significant relationships to the Blacky were established. Had Charen employed the Blacky correctly, there would still be no reason to expect congruence with Rorschach findings. Berlow's results are not introduced to set off a debate on the relative merits of the two instruments, but simply to point out that there is no place for the Rorschach in Blacky's orbit of construct validation.

Finally it is important to keep in mind that even when the technique is administered and utilized correctly, the retest method, reflecting as it does an unknown mixture of changes in set, familiarity with items, and perhaps personality changes as well, cannot properly be employed to study reliability of responses to the Blacky Pictures.

Received March 23, 1956.
Early Publication by the Editor.

References

1. Aronson, M. A study of the Freudian theory of paranoia by means of the Blacky Pictures. *J. proj. Tech.*, 1953, 17, 3–19.
2. Berlow, N. Psychosexual indicators on the Rorschach test. Unpublished doctor's dissertation, Univer. of Michigan, 1952.
3. Blum, G. *The Blacky Pictures.* New York: Psychological Corp., 1950.
4. Blum, G. Revised scoring system for research use of the Blacky Pictures. Ann Arbor, Michigan: Author, 1951 (mimeographed).
5. Charen, S. Reliability of the Blacky Test. *J. consult. Psychol.*, 1956, 20, 16.

Journal of Consulting Psychology
Vol. 20, No. 5, 1956

A Reply to Blum

Sol Charen
Montgomery County Mental Health Clinic, Rockville, Maryland

Blum pinpoints emphasis on my title to suggest that the technique I used is incorrect. A title is only intended to furnish brief information about an article. Mine could have been clearer. It is also necessary to read the article to find out what it is about and not rely solely on the title. Clearly stated throughout my brief report, "The Reliability of the Blacky Test," is the fact that it deals with the low reliability of the multiple-choice questions of the Blacky test, as found in my doctoral research (2). This is all my article deals with, not the total Blacky test, and Blum is unwittingly in accord with my text when he states in his reply ". . . multiple-choice items from the inquiry, . . . [are] closest to consciousness and, therefore, most easily distorted by the subject." As he points out, though, the multiple-choice questions do furnish one of the four sources which contribute to each dimensional score. The question of reliability of the total Blacky test is therefore involved when a major score source may be questionable.

Since spontaneous stories were not obtained from my chronically ill patient subjects I did not follow his technique. How this omission affected their responses to the multiple questions for each cartoon I do not know. But Blum is on shaky ground when he implies that this omission somehow affected the reliability since he has no reliability studies of his own for either the total Blacky test or any of its four scoring sources. Without such data, his assumption is purely hypothetical that the omission of the spontaneous stories somehow affected the reliability in the test-retest situation.

Blum has lifted a sentence out of context. Nowhere did I use the Rorschach test to validate either the Blacky test or its multiple-choice items. The Rorschach and fifteen paper-and-pencil tests were used to evaluate whether my patients in a four-month period showed test evidence of basic personality changes as a result of chronic illness and extensive hospitalization. No test evidence of such changes was found. Only the multiple-choice items of the Blacky test indicated that some changes took place in my experimental subjects. These, in general, suggested they were nonregressed during illness testing and were regressed when retested shortly before discharge from the hospital. I believe I am consequently justified in assuming that the equivocal results found in the Blacky multiple-choice items may have been due to the factor of the low reliability which was found to exist.

One final comment about the last paragraph of Blum's reply is in order. The question of set should not be involved in the Blacky test as it might be in other personality tests since his scoring allows for evasiveness or attempts to use defenses. Familiarity with items is not likely to be a problem when a four-month or longer period elapses between the test and retest. Finally, since the Blacky test purports to measure, "the deeper recesses of personality" (1), namely the genetic psychosexual dimensions which should not change in the individual unless psychoanalysis or a marked life experience occurs, personality change is also not a factor which should prevent use of the retest method.

Received June 25, 1956.
Early publication by the editor.

References

1. Blum, G. S., & Hunt, H. F. The validity of the Blacky pictures. *Psychol. Bull.*, 1952, **49**, 238–250.
2. Charen, S. Regressive behavior changes in the tuberculous patient. *J. Psychol.*, 1956, **41**, 273–289.

PSYCHOLOGICAL TEST REVIEWS

Doll, Edgar A. *Vineland Social Maturity Scale.* Minneapolis: Educational Test Bureau, 1947.

Reviewed by

Morris Krugman

Board of Education, New York, N. Y.

An instrument that has been in use for thirty years is hardly news. The first article about the *Vineland Social Maturity Scale* by Edgar Doll appeared in 1935, after almost ten years of use and development at the Vineland Training School; the last revision was made in 1936, and the manual accompanying the current form of the scale was published in 1947. Even a casual inspection of psychological literature reveals more than a hundred titles discussing or mentioning use of the scale, and many more are undoubtedly available.

Something new, however, has been added, with the publication of *The Measurement of Social Competence* by Edgar A. Doll in 1953. For the first time in thirty years, the rationale, the details of the standardization process, scoring samples, methods of interpretation, and other pertinent data usually supplied when psychological instruments are first presented for general use, are made available for the *Vineland Social Maturity Scale.* Although the brief 1947 manual mentions the care with which the standardization process was carried out, any real idea of the meticulousness of this process can be gained only from the voluminous 1953 manual. Each of the 117 items of the scale is described in minute detail with possible variations, normative curves are presented for each item, and means, medians, *SD*s and *CR*s supplied. Line drawings illustrating typical behavior are added. One wonders whether the treatment of data was overelaborated, since the norms are based on ten males and ten females at each age level from 0–1 to 30, a total of 620 subjects.

The scale is too well known to require detailed description here. Briefly, it consists of 117 items, scaled in order of difficulty and arranged by years from 0 to 30. The items are distributed among six categories: self-help (general, eating, and dressing), self-direction, occupation, communication, locomotion, and socialization. "Each item is conceived as representing a general growth in social responsibility which is expressed in some detailed performance as an overt expression of that responsibility. Consequently, the value of the detailed items is to be determined principally by the extent to which they reflect this personal independence in personal activities, in respect to which the detailed performances are otherwise relatively unimportant." The items "reflect progressive freedom from need of assistance, direction, or supervision on the part of others."

Although direct rating by the examiner is not ruled out if opportunity for observing or knowing the subject is available, the usual method of use is by interviewing an informant who is thoroughly familiar with the behavior of the subject. The fact that the subject being examined and rated need not be present or observed is both an advantage and a disadvantage. The obvious advantage is the case of applying the scale by questioning a generally mature, competent adult about an immature, incompetent, or handicapped individual. The disadvantage lies in the possibility of gross distortion and incorrect rating. The reviewer has always experienced a feeling of greater confidence about ratings when the subject was seen even briefly after the usual rating by means of an informer, and when a few items of the scale have been spot checked.

Usually, the publication of basic data about a clinical instrument serves to clarify issues and strengthen the use of the instrument. This is not the case with the *Vineland Social Maturity Scale.* That the scale is a very useful clinical adjunct with many different types of subjects is attested to by the almost uniformly favorable literature over the past twenty years. This reviewer has used it for many years and has tremendous respect for it. But this respect is not increased by careful reading of the new manual. The small number of cases at each level has been mentioned and indeed was known from the earlier manual, but the elaborate statistical and graphic treatment of each item is not convincing, since, in spite of the fact that continued appraisal of each item was made statistically, experimentally, and clinically, not one of the 117 items has been shifted up or down, nor have items been replaced or modified. Equally questionable is the upper end of the scale, intended for normal and superior adolescents and adults. Although based on normative data, this part of the scale could well be an extrapolation of the lower end of the scale, and it hardly makes sense. An ingenious method of measuring defects in

social maturation of children and adults, and indicating the level of social development of children who are not necessarily retarded in development, the Vineland Scale is artificial when applied to normal or superior adolescents or adults. Most users of the scale simply ignore the upper levels.

The *Vineland Social Maturity Scale* has its greatest usefulness in distinguishing between real and pseudo-feeblemindedness, in providing a measure of social development in studies of child growth and development, and in appraising levels of social function among handicapped individuals of many kinds and in many types of pathology, particularly in cases of arrested development or deterioration. It has made a permanent place for itself in clinic and laboratory, and newer approaches to the appraisal of social maturity have not succeeded in replacing it. The fact that it has not been found necessary to modify or "improve" the scale over these many years can be accepted as evidence of the carefulness with which it was originally constructed, and the keen clinical insight that obviously determined the original construction and standardization.

Doll, Edgar A. *Vineland Social Maturity Scale.* Minneapolis: Educational Test Bureau, 1947.

Reviewed by

Edwin S. Shneidman

VA Neuropsychiatric Hospital, Los Angeles

The *Vineland Social Maturity Scale* is not a psychological test in the usual sense, nor does this review pretend to be the usual type of test review. Most tests are samples of modes of reacting, perceiving, or adjusting, or are extrapolations from segments of behavior. The *Vineland Social Maturity Scale,* on the other hand, attempts to measure behavior directly. Although Doll says that *"The scale is not a rating scale and scores are not to be based on mere opinions"* (italics his), nevertheless the scale would seem to be in the same tradition and of the same genre as the Gesell *Development Schedules,* the Malamud and Sands *Psychological Rating Scale,* the Ferguson-McReynolds *Hospital Rating Scale,* etc. The *Vineland Social Maturity Scale* is a well known (to more than one generation of clinicians) measure of the social and adaptational skills, habits, traits, coping behaviors, and interpersonal reactions of subjects between the ages of 0–1 and 25 plus. Doll indicates that the purpose of his scale is "To provide an over-all measure of personal adequacy in terms of social maturity." He further indicates that the scale is a "descriptive analysis of maturation in social competence, as a standardized schedule of behavior development in terms of its habitual expression." He feels that it has use in establishing a standard schedule of normal development, in measuring deviations which may be significant in such problems as delinquency, amentia, adoption of children, etc., in measuring improvement following training or therapy, etc.

There are 117 items in the scale, arranged in order of average age norms. The items have been chosen to report some aspect of the individual's ability to look after his own practical needs. Social quotients (SQ) "generally comparable to Binet MA and IQ can be obtained from the sum of the items scores." The items themselves range from such as "'Crows'; laughs" (0–1) to "Advances general welfare" (XXV). The items at the upper end of the scale (ages 20 up) such as "Uses money providently," "Contributes to social welfare," "Systematizes his own work," "Inspires confidence," "Promotes civic progress," "Advances the general welfare," etc. seem extremely hard to evaluate. Nor does Doll's injunction that *"The informant does not make the scoring judgment. This is done by the examiner after obtaining from the informant as much detail as practicable regarding the behvaioristic facts which reveal the manner and the extent of the subject's actual performance on each item"* (italics his) really circumvent the essential problem of observing and rating items that are essentially difficult to rate.

The main purpose of this summary is neither to describe the *Vineland Social Maturity Scale* nor to present a survey of the published research on it; it is rather to discuss some aspects of its implied rationale, and some cautions in its application. Certainly the careful use of the scale to distinguish those subjects who do deviate from the norms of "social maturity," so that further study of those individuals can be made, would be a useful and rewarding application of this instrument. However—and this is the burden of this review—the *Vineland Social Maturity Scale* lends itself easily to other interpretations.

Ordinarily, when one uses an intelligence test in order to obtain an IQ, the main purpose of such an effort is to ascertain what the IQ is, although the implicit assumption may be that the higher the IQ, the better. This implicit assumption operates in a much more subtle and in a more questionable way in the case of the *Vineland Social Maturity Scale.* Specifically, the assumption is that social behavior (including social motivation) is desirable as it approaches social conformity. Further, the implicit assumption is that optimal behavior lies in adjustment to a standardized society and conformity to the "normal." Without becoming embroiled in the intricacies of the late Robert Lindner's concepts of rebellion and salutary nonconformity, it still remains that the *Vineland Social Maturity Scale* deemphasizes the individual's fealty to his own image, his strength to be himself, his positive (that is creative or effective) idiosyncracies, his areas of adaptation which may bring him pleasure, or the application of his resources (within his total personality organization) toward creativity or self-actualization. The recent work of Jean MacFarlane at Berkeley and Lois Murphy at Sarah Lawrence would seem to indicate that the conforming person is not necessarily the best adjusted person. One is reminded of a section of D. B. Klein's book *Mental Hygiene* entitled, "In Defense of the Shy Child" (who does not

necessarily become schizophrenic) and one wonders if a section should be added in defense of the nonconforming child (who need not necessarily become maladjusted, psychopathic, or feebleminded).

The point of all the above statements is that the *uncritical* use of the *Vineland Social Maturity Scale* may imply an unwarranted assumption of the merits of conformity and standardization and to permit a confusion between aberrant or pathological nonconformity (schizophrenia, for example, at one extreme). To make these assumptions would be to gainsay the infinite variety and the rich diversification of talents and gifts of human subjects.

As a clinical psychologist—and as a father of four—I am not sure that I know what the "best" age for toilet training, for example, really *is*. I would also ask whether the *child* has a problem if he isn't toilet trained by "that" age or rather cannot the parent create problems by imparting parental and societal anxieties and concerns—*his* problems—to the child.

The major reference for the *Vineland Social Maturity Scale* is Doll's *The Measurement of Social Competence* (Minneapolis: Educational Test Bureau, 1953). Two shorter references are the *Manual of Instructions* (published by the Educational Test Bureau, dated 1947) which cites 59 references, and Dr. Doll's chapter (in the second part of Chapter 20) on his scale in Weider's *Contributions Toward Medical Psychology* (New York: Ronald Press, 1953).

Rotter, Julian B., & Rafferty, Janet E. *The Rotter Incomplete Sentences Blank (College Form)*. New York: Psychological Corporation, 1950.

Reviewed by

Louis Long

The City College of New York

The Rotter Incomplete Sentences Blank (ISB) "is a semistructured projective technique in which the subject is asked to finish a sentence for which the first word or words are supplied" (Manual, p. 3). Historically, the incomplete sentence method is closely related to word association tests, and the ISB has a long history of evolution. Sentence forms used by Rotter and Willerman, Shor, Hunt, Holzberg, Rohde and Hildreth, and Payne have had an effect upon the ISB. As in all projective situations it is assumed that the subject in completing the sentences will reflect something of his own personality—his fears, desires, wishes, anxieties.

The ISB consists of 40 "stems" that are to be completed by the subject. The completions are then scored by comparing them with responses that have been assigned a scale value varying from 0 to 6. The total score is an index of maladjustment.

The ISB may be used in two very different ways: (*a*) as a gross screening instrument; and (*b*), as a clinical tool. It seems desirable to evaluate these two uses separately.

Evaluation of the ISB as a screening instrument. Biserial coefficients of 0.50 and 0.62 between scores and classifications of adjusted and maladjusted students are presented in the manual to show how effective the test is when used as a screening instrument. The small number of cases upon which these data are based should, however, warn clinicians to proceed with caution in using the ISB for this purpose. Each person using the test will do well to study the effectiveness of it in his own local situation. This in essence means that the ISB must be validated locally which will necessarily require setting up criterion groups of adjusted and maladjusted individuals. In their manual, Rotter and Rafferty describe in detail the method they used to establish their criterion groups. It is easy to sense their dissatisfaction with the use of ratings of adjustment or maladjustment as supplied by instructors teaching classes in effective study and in mental hygiene, or by advanced student clinicians. In the reviewer's opinion, the difficulty of establishing satisfactory criterion groups will inevitably make it impractical for many colleges to use the ISB. If, however, the aggressive research program which seems to be underway at present is continued it may eventually be possible for colleges to use the ISB as an aid in screening students for possible referral to a counseling agency.

Evaluation of ISB as a clinical tool. To facilitate the use of the ISB as a clinical instrument the authors of the Manual have explained their scoring procedure in detail, have given many examples of the various kinds of responses, have presented practice cases along with the correct scoring of the responses, and have given a clinical interpretation of six records. Each of the six cases begins with a brief background sketch of the individual, followed by a record of his responses on the ISB, and an analysis of the case in terms of familial attitudes, social and sexual attitudes, general attitudes, and character traits. The reports conclude with a summary, and a discussion of the contribution to the ISB. These cases are presented to illustrate the potential value of the ISB and it is emphasized by the authors that only *some* of the possible interpretations are included in the analysis. As one reads the analyses and studies the responses other interpretations often seem plausible. This in a sense is an indication of the value of the test since one of the main purposes of a projective situation is to furnish leads that may be explored in an effort to gain clinical insights. The authors are effective in showing how the ISB adds to their understanding of clients. It would be interesting to know how the ISB's contribution compares with that of the Rorschach or the TAT. Specific information about the relative contribution of the several projective devices is urgently needed.

The authors of the manual are careful to point out the limitations of the ISB and to remind readers that all data presented relate to the college form. It is somewhat unfortunate that high school and adult forms are listed as being available. Until such time as manuals and norms are available it would seem desirable to keep such listings out of test catalogues.

INTERPERSONAL DIAGNOSIS OF PERSONALITY*

- MULTILEVEL CLINICAL DIAGNOSIS AND PROGNOSIS
- ANALYSIS OF GROUP DYNAMICS IN PSYCHOTHERAPY OR INDUSTRIAL MANAGEMENT
- FAMILY DIAGNOSIS IN CHILD GUIDANCE AND MARITAL COUNSELING
- MEASUREMENT OF CHANGE IN PSYCHOTHERAPY
- PERSONALITY RESEARCH IN ANY OF THESE AREAS

Test Materials Now Available

INTERPERSONAL CHECK LIST. A 128 item test by means of which the subject diagnoses himself, family members, fellow group members, therapist, etc. Also used by others to rate the subject.

RECORD BOOKLET FOR INTERPERSONAL DIAGNOSIS OF PERSONALITY. Clinical diagnosis and prediction of interpersonal behavior.

RECORD BOOKLET FOR ANALYSIS OF GROUP DYNAMICS. Interpersonal perceptions by each member in a group of every other member. Diagnosis of group dynamics—overt and covert patterns of relationships.

RECORD BOOKLET FOR FAMILY DIAGNOSIS. Multilevel diagnosis of husband and wife, their misperceptions of each other and of their children. Oedipal structure of the marriage.

MANUAL FOR INTERPERSONAL DIAGNOSIS

The manual provides a step-by-step description of all the procedures for interpersonal diagnosis. Administration and scoring of each interpersonal test is outlined. All steps involved in standardization, plotting of scores, interpersonal diagnosis and the calculation of conflict indices are discussed and illustrated.

The use of the interpersonal system in research is given detailed consideration. Particular attention is given to the handling of multilevel data in personality research.

* Developed by Leary et al at the Kaiser Foundation, Oakland, California. The Interpersonal System has been used in its prepublication form by over 40 institutions. See:

Interpersonal Diagnosis of Personality, Ronald Press, 1956.
J. Pers., 1951, 20, p. 144; 1954, 23, p. 129; 1955, 24, p. 96; 1956, Dec.
Gp. Psychother., 1954, 7, p. 7.
J. Abnorm. Soc. Psychol., 1955, 50, p. 110.
Psychiat., 1955, 18, p. 147.
J. Clin. Psychol., 1956, 12, p. 133.

A descriptive catalog and sample copies of these materials can be obtained from:

PSYCHOLOGICAL CONSULTATION SERVICE
1230 Queens Road Berkeley 8, California

THE DEVEREUX FOUNDATION

Since 1940, non-profit sponsor of

DEVEREUX SCHOOLS

Established 1912

THERAPEUTIC SCHOOLS

VOCATIONAL COMMUNITIES

REMEDIAL CAMPS

SCHOOLS

EARL D. BOND DIVISION — Pennsylvania
WALTER L. TREADWAY DIVISION — California

A complete scholastic program, in residence, for boys and girls presenting emotional difficulties, who can benefit by a program of therapeutic education, supplemented by psychoanalytic therapy, where indicated.

Separate facilities and campuses for different age groups, from pre-kindergarten through junior college.

COMMUNITIES

LEO KANNER DIVISION — Pennsylvania
RANGE DIVISION — California

"Life-experience" and vocational programs for children, adolescents, and young adults with impaired intellectual or CNS function.

Separate, self-contained campuses for homogeneous groups, in terms of age and social maturity.

CAMPS

PENNSYLVANIA — MAINE — CALIFORNIA

Each Devereux School or Community has its own individual camp. Remedial academic programs, full recreational facilities, permit therapy to continue through the long summer holiday.

Professional inquiries will be welcomed by JOHN M. BARCLAY, *Director of Development*
Devereux Foundation, Devon, Pennsylvania

"HUMANI NIHIL A' ME ALIENUM PUTO"
THE DEVEREUX FOUNDATION
CALIFORNIA — PENNSYLVANIA

Devereux Schools

UNDER THE DEVEREUX FOUNDATION

HELENA T. DEVEREUX, *Director*

Professional Associate Directors

ROBERT L. BRIDGEN, PH.D. — EDWARD L. FRENCH, PH.D.
MICHAEL B. DUNN, PH.D. — ROBERT T. GRATTAN, M.D.
J. CLIFFORD SCOTT, M.D.

Santa Barbara, California — **Devon, Pennsylvania**